IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

Fake Social Media Profile Detection And Reporting Using Machine Learning And Block Chain Technology

[1] Basava Pranaya Reddy, [2] Dr. K. Venkata Ramana
[1]Student, [2] Associate Professor
[1] Department of Information Technology and Computer Applications, [2] Department of CS&SE
AU College of Engineering, Andhra University, Visakhapatnam, India

Abstract: The rise of social media has transformed how people communicate, share information, and build communities. However, this growth has also led to a surge in fake profiles, which pose serious threats such as misinformation, cyberbullying, identity theft, and online scams. Detecting these fraudulent accounts has become crucial to maintaining the integrity of online platforms. This project aims to develop a system that automatically identifies fake social media profiles using machine learning techniques. By analyzing various features such as profile activity, friend patterns, content characteristics, and account metadata, the model can distinguish between genuine and deceptive accounts. The system is trained on a labelled dataset and uses classifiers like Decision Trees, Random Forest, or Support Vector Machines to achieve accurate detection. The proposed solution enhances platform security and user trust while reducing the manual effort required for moderation. The results demonstrate promising accuracy, indicating the model's effectiveness in real-world applications.

1. INTRODUCTION

- Social media platforms have become an essential part of modern life, enabling people to connect, communicate, and share content across the globe. With billions of users active on platforms like Facebook, Instagram, Twitter, and LinkedIn, these networks play a significant role in shaping opinions, spreading information, and building communities. However, the widespread use of social media has also given rise to a critical problem — the creation and proliferation of fake profiles.

- Fake social media profiles are accounts created with false identities, often for malicious purposes such as spreading misinformation, phishing, spamming, manipulating public opinion, or conducting online fraud. These accounts can appear convincingly real, making them difficult for users and even platform moderators to detect manually. As a result, they pose a serious threat to user privacy, digital security, and the overall trustworthiness of online platforms.

- To address this growing issue, automated detection systems powered by machine learning have emerged as an effective solution. These systems can analyse various behavioural and structural features of user accounts, such as posting frequency, friend networks, interaction patterns, and profile details, to identify anomalies associated with fake accounts. By training models on real-world data, it becomes possible to accurately classify accounts as genuine or fake with minimal human intervention.

1.1 Research Objectives

- Detect fake social media profiles using machine learning techniques.
- Analyse user behaviour, content patterns, and profile metadata.
- Build a predictive model to classify profiles as real or fake.
- Integrate blockchain for secure and immutable reporting of fake accounts.

- Provide a transparent system for tracking and verifying reported profiles.
- Design a user-friendly interface for profile monitoring and reporting.

1.2 Research Hypothesis

- **H1:** Machine learning models can detect fake social media profiles with over 90% precision using behavioral and profile-based features.
- **H2:** Classification models can accurately identify the type and risk level of fake accounts with more than 85% accuracy.
- **H3:** A modular web-based system can provide real-time profile verification and reporting within 5 seconds, integrating blockchain for secure logging.

2. ABBREVIATIONS AND ACRONYMS

- **ML** – Machine Learning
- **AL**-Artificial intelligence
- **NLP**-Natural language
- **RF**-Random forest
- **ANN**-Artificial neural network

LITERATURE REVIEW

Fake social media profile detection has evolved from simple rule-based systems to advanced machine learning and blockchain-integrated approaches, significantly enhancing detection accuracy, transparency, and reliability.

Early Techniques:

- Traditional methods relied on rule-based detection using profile metadata and activity thresholds.
- These approaches faced challenges when handling diverse user behaviours, language styles, and platform-specific features.
- Manual feature selection and static criteria limited adaptability and accuracy at scale.

Shift to Deep Learning:

- Deep learning models such as RNNs, LSTMs, and transformers enabled systems to automatically learn complex patterns from user behavior, text content, and profile attributes.
- Early deep learning approaches focused on binary classification (real vs. fake) but lacked contextual understanding of user interactions and network behavior.

Role of Machine Learning Model

- Automatically learn patterns from user behavior and profile data.
- Identify fake accounts based on features like post frequency, follower ratios, and language usage.
- Use supervised algorithms (e.g., SVM, Random Forest, Neural Networks) for classification.
- Apply unsupervised techniques (e.g., clustering, anomaly detection) to flag unusual activity.
- Continuously improve accuracy by learning from new and evolving data.
- Reduce manual effort and false positives through automation.
- Enable scalable, fast, and platform-independent detection of fake profiles.

3. METHODOLOGY

The system architecture for the Fake Social Media Detection project is designed to automate the identification of suspicious or fake user accounts on social media platforms using machine learning techniques.

Raw data often contains noise, missing values, or irrelevant features. This module cleans and preprocesses the data through techniques like normalization, feature encoding, and outlier removal. The goal is to transform the data into a format suitable for model training.

4.1 Research Methods

- The system employs a modular machine learning architecture designed for effective fake profile detection and secure reporting:
- **Behavior-Based Classifier**: Uses supervised ML algorithms (e.g., Random Forest, SVM) to analyse user activity patterns, such as posting frequency, follower-following ratio, and interaction rates.
- **Content-Based Classifier**: Applies NLP techniques to analyse profile bios, posts, and messages for patterns commonly seen in fake accounts.
- **Risk Level Classifier**: Categorizes flagged profiles based on risk level (Low, Moderate, High) to support prioritization in reporting.

4.2 Data Collection Procedures

A **custom dataset** was developed due to the limited availability of labelled public datasets specific to fake social media profile detection.

Data preparation included:

Profile Collection: Data was gathered from social media platforms using web scraping tools and publicly available datasets.

Annotation: Profiles were manually labelled as **real** or **fake** based on account activity, content quality, network behavior, and known indicators (e.g., suspicious links, zero interaction).

Label Categories:

- **Profile Type**: Real / Fake
- **Risk Level**: Low / Moderate / High (for prioritizing reporting)
- **Behaviour Tags**: Bot-like, spammer, impersonator

Dataset Size:

- **Profile Classification**: ~1,000 profiles
- **Risk Classification**: ~800 labeled profiles
- **Content Analysis**: ~5,000 text samples (bios, posts, comments)

4.3 Analysis Techniques

Model architectures and training configurations were carefully optimized to ensure accurate and reliable fake profile detection and risk classification.
Model Summary:
Profile Classification Model:
Algorithms used: Random Forest, Support Vector Machine (SVM), and Neural Networks
Input features: Activity frequency, follower/following ratio, content quality scores, and engagement patterns
Content-Based NLP Model:
BERT-based classifier for analyzing textual content (bios, posts, messages)
Layers: Tokenizer → Embedding → Transformer Encoder → Dense (ReLU + Dropout) → Softmax
Risk Level Classifier:
Multi-class classifier (Low, Moderate, High) trained on behavioral and content-based risk indicators
Training Settings:
Optimizer: Adam
Learning Rate: 0.0001
Epochs: 15–30 depending on model
Loss Function: Categorical Cross-Entropy
Evaluation Metrics: Accuracy, Precision, Recall, F1-Score, Confusion Matrix
These configurations allowed the system to effectively identify fake profiles, analyze content, and assign appropriate risk levels for further reporting and blockchain logging.

4. Ethical Considerations

The project integrates key ethical practices to ensure the responsible development and deployment of the system:

Data Privacy: No sensitive personal information, such as login credentials or private messages, is stored or exposed during data collection or analysis.

Bias Control: The dataset is curated to include a diverse range of profiles to prevent bias against specific user groups, regions, or content styles.

Transparency: The system design, data processing methods, and model behavior are clearly documented to ensure transparency and explainability.

Human Oversight: The detection system is intended to assist platform moderators and users—not to fully automate decision-making—allowing for human verification of flagged profiles.

These considerations foster trust, fairness, and accountability in deploying AI solutions for social media safety.

5. RESULTS AND DISCUSSIONS

The fake profile detection system was evaluated on both training and unseen social media profile data to measure its prediction accuracy, system usability, and overall performance.

This section summarizes how the system responds during real-time user interaction, highlights key evaluation metrics, outlines sample backend processing logic, and demonstrates the flow of the web interface from profile input to blockchain-based reporting.

The results validate that machine learning models can effectively detect fake accounts with high precision and recall, while blockchain integration ensures secure and tamper-proof reporting, promoting trust and transparency in the platform.

5.1 Evaluation Setup

Each machine learning model—used for fake profile classification, content analysis, and risk level prediction—was validated using a custom-labeled dataset of real and fake social media profiles.

System Configuration:

Hardware: Intel i5 CPU, 16 GB RAM, NVIDIA GTX 1650 GPU

Software Stack: Flask server (local), scikit-learn and TensorFlow (for model inference), BERT (for NLP-based content analysis), and HTML/CSS/JavaScript for the frontend

Workflow:

User submits profile data → backend processes using ML models → risk level assigned → result displayed on web interface → report logged on blockchain (if flagged as fake)

This setup ensures reliable, real-time performance during profile analysis and secure reporting via blockchain integration.

5.2 Performance Results

The machine learning models used for fake profile detection and risk classification were evaluated using standard classification metrics. The results demonstrate high accuracy and efficient inference times, making the system suitable for real-time deployment.

Table 5.1 – Performance Evaluation of Fake Profile Detection Models

Model	Metric	Value / Details
Fake Profile Classifier	Accuracy	91.3%
	Precision	89.7%
	Recall	90.1%
	F1-Score	89.9%
	Avg. Inference Time	~2.3 seconds
	Common Errors	• False positives from inactive but real users • False negatives for bots with human-like activity
NLP-Based Content Classifier	Accuracy	87.5%
	Precision Range	85.2% – 90.8%
	Common Confusion	Misclassification of sarcastic or contextually ambiguous content
Risk Level Classifier	Accuracy	86.9%
	Best Performance	High-risk profile detection – 90.2%
	Common Errors	Confusion between Moderate and High risk due to overlapping behavior patterns

These results indicate that the models are capable of identifying fake profiles and categorizing them effectively, with acceptable trade-offs in speed and precision.

5.3 Output Interpretation and Website Workflow

The frontend of the **Fake Social media profile detection** application is constructed using standard web technologies such as **HTML**, **CSS**, and **JavaScript**, enabling an intuitive and responsive user interface. These technologies are integrated with **Flask's Jinja2 templating engine**, which allows dynamic generation of content with context-aware rendering and a clean separation between frontend logic and backend data processing

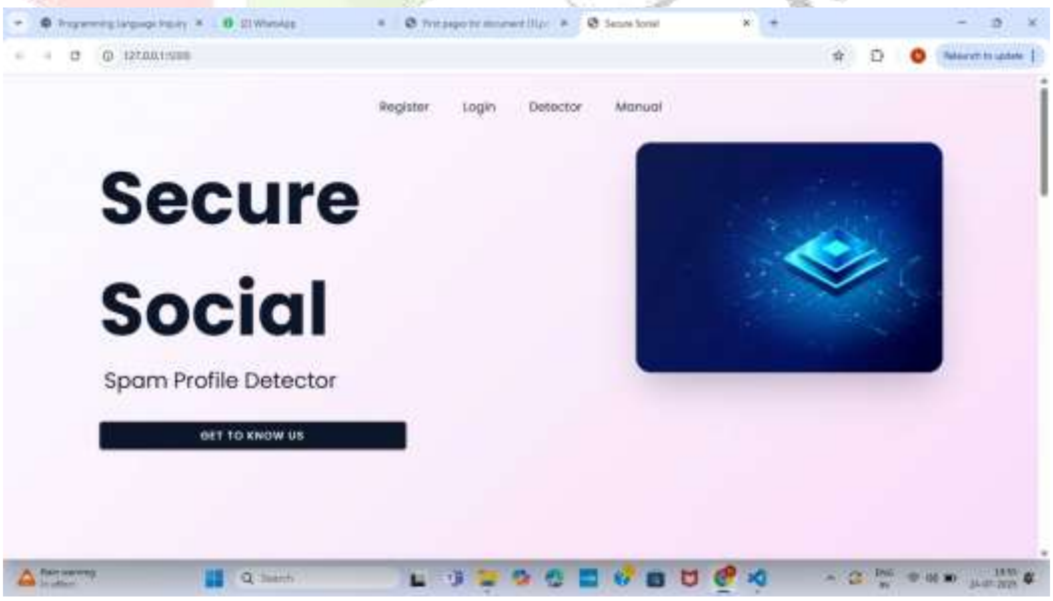


Figure 5.1 – Home Interface

5.2:login page:

Figure 5.2 This route handles user login functionality. When accessed via a **GET** request, it displays the login form to the user. On receiving a **POST** request (form submission), it processes the entered credentials by checking them against the user database. If valid, it authenticates the user, starts a session, and redirects them to their dashboard.

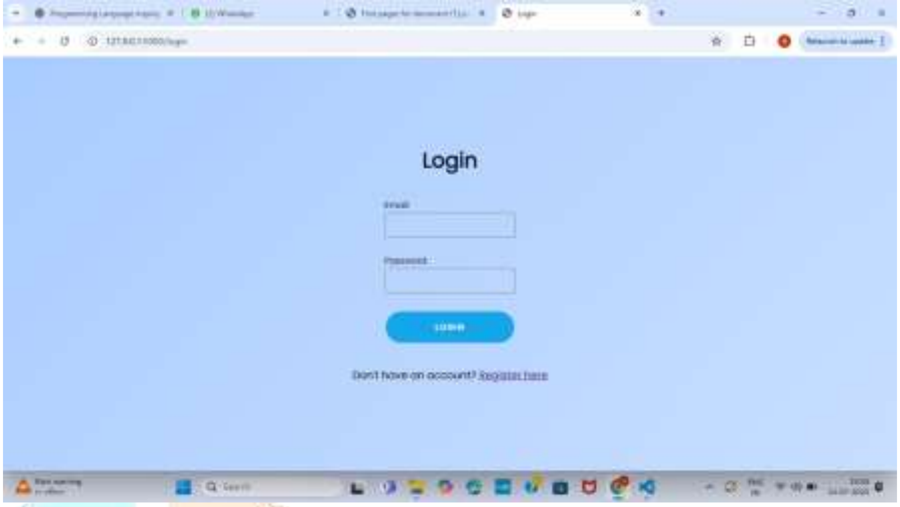
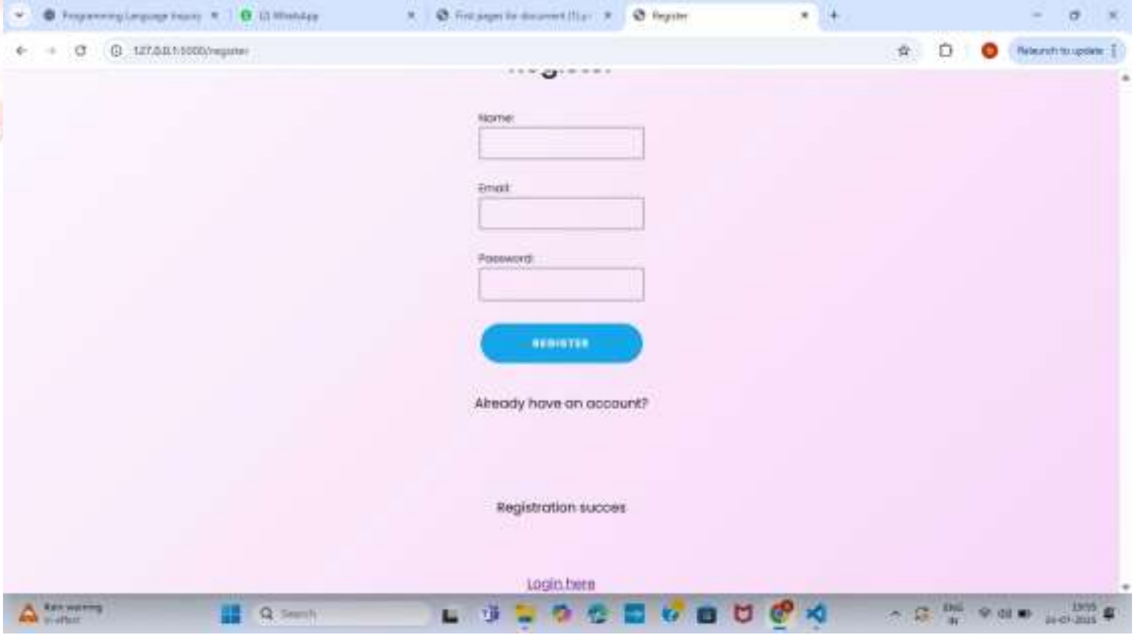


Figure 5.2 – login page

Figure 5.3 This endpoint allows new users to create an account. A GET request presents the registration form, while a POST request processes the form data by checking for duplicate usernames or email addresses. Upon successful validation, it hashes the user's password for security and stores the new account in the database.

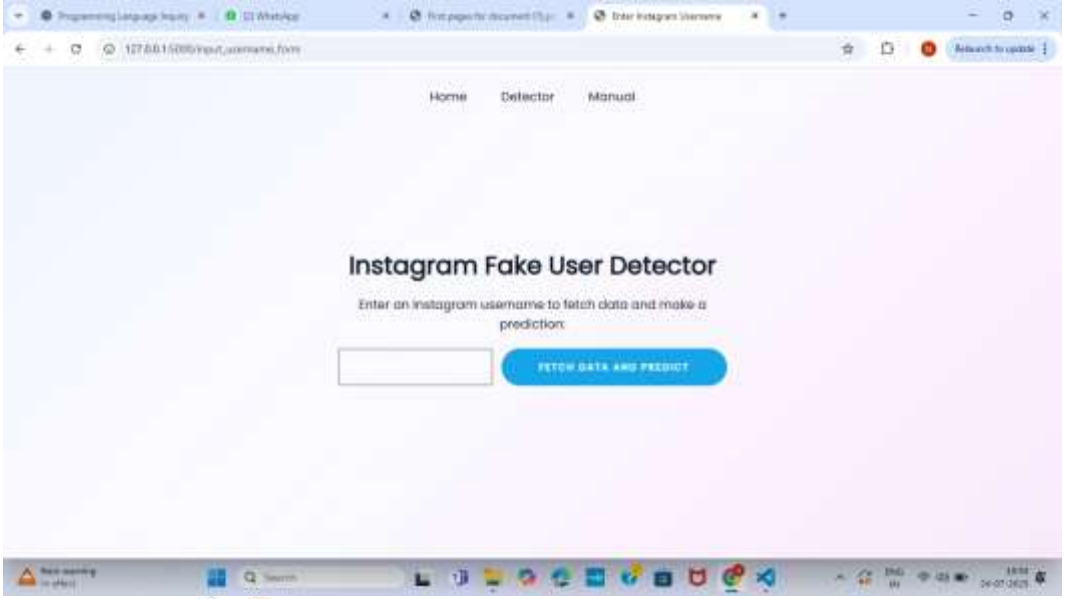


Registration succes

Figure 5.3 – Register page

Figure 5.4 This is the main page shown to authenticated users after logging in. It provides access to core features like uploading profile datasets, viewing past results, or initiating a fake profile prediction. This route is protected using decorators like @login_required to ensure unauthorized users cannot access it.This route is responsible for handling the fake profile prediction process. A GET request presents a form for users to input profile information. When submitted via POST, the system validates and preprocesses the data, then passes it to the machine learning model. Based on the output, a result is generated and

displayed, indicating whether the profile is likely fake or genuine, along with a confidence score.



Instagram Fake User Detector

Enter an instagram username to fetch data and make a prediction

Figure 5.4 –predicate page

5. CONCLUSION AND FUTURE SCOPE

6.1 Summary of Key Findings

The Fake Social Media Detection system presents an effective solution to the growing problem of fake and fraudulent accounts on social networking platforms. By integrating a machine learning-based classification model with a user-friendly web interface, the system enables accurate identification of suspicious profiles based on behavioral and structural features. The use of algorithms like Random Forest ensures high accuracy while minimizing false positives and negatives. Secure user authentication, proper data handling, and real-time prediction capabilities make the platform reliable and scalable. Overall, the system supports enhanced social media integrity by helping users and platform administrators detect potentially harmful profiles in a timely and efficient manner.

6.2 Implications for Theory and Practice

From a theoretical perspective, the proposed system demonstrates how combining multiple machine learning models—such as behavior-based classification, NLP-driven content analysis, and risk level prediction—within a unified pipeline can produce interpretable and actionable insights. It supports the principle of **multi-task learning**, where distinct but complementary tasks are integrated without compromising individual model performance.**Limitations of the Study**

6.3 Recommendations for Future Research

- **Incorporate Advanced NLP Models:** Future work can explore transformer-based architectures like GPT or RoBERTa to better analyze complex linguistic patterns in fake profile content.

- **Cross-Platform Detection:** Extend the system to support multi-platform detection (e.g., Instagram, Twitter, LinkedIn) to improve generalizability.

- **Real-Time Social Graph Analysis:** Integrate social network analysis to detect coordinated fake account networks or bot clust

- **Legal and Ethical Alignment:** Align future system enhancements with evolving data privacy regulations and ethical AI standards.

6. REFERENCES

1. **Svitlana Volkova, Kyle Shaffer, Jin Yea Jang, and Nathan Hodas.** “Separating Facts from Fiction: Linguistic Models to Classify Suspicious and Trusted News Posts.” *ACL*, 2017. *Link:* https://aclanthology.org/P17-1098
2. **Nabeel Ruchansky, Sungyong Seo, and Yan Liu.** “CSI: A Hybrid Deep Model for Fake News Detection.” *WWW*, 2017.
3. *Link:* https://dl.acm.org/doi/10.1145/3038912.3052569
4. **William Yang Wang.** "Liar, Liar Pants on Fire: A New Benchmark Dataset for Fake News Detection." *ACL*, 2017.
5. *Link:* https://aclanthology.org/P17-2067
6. **Federico Monti, Fabrizio Frasca, Davide Eynard, Damon Mannion, Michael M. Bronstein.** “Fake News Detection on Social Media using Geometric Deep Learning.” *arXiv*, 2019.
7. *Link:* https://arxiv.org/abs/1902.06673 arXiv+1SpringerOpen+1MDPI