ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 7| July 2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

AI BASED FAKE PROFILE DETECTION SYSTEM FOR INSTAGRAM USING MACHINE LEARNING

A complete study of AI based fake profile detection — architecture, training, evaluation, and real-world applications

G. Vennela Vijaya Lakshmi[1], V. Gulabi[2], Y. Shoba[3], D. Aruna Padma[4], V. Sailaja[5], K. Leela Pavani[6], Sunil B[7], B. Jyothi[8]

[1]*M.Sc. Computer Science, Visakha Govt Degree & PG College For Women*
[2]*M.Sc. Computer Science, Visakha Govt Degree & PG College For Women*
[3]*M.Sc. Computer Science, Visakha Govt Degree & PG College For Women*
[4]*Lecturer in Computer Science,Visakha Govt Degree & PG College For Women*
[5]*Lecturer in Computer Science,Visakha Govt Degree & PG College For Women*
[6]*Lecturer in Computer Science,Visakha Govt Degree & PG College For Women*
[7]*Lecturer in Computer Science,Visakha Govt Degree & PG College For Women*
[8]*Lecturer in Computer Science,Visakha Govt Degree & PG College For Women*

ABSTRACT

Social media has completely changed the way we communicate, connect with others, and share our lives online. Among the many platforms available today, Instagram stands out as one of the most widely used, with millions of people sharing photos, videos, and personal moments every day. But this popularity comes with a downside – the platform has also seen a sharp rise in fake profiles, which pose real risks to user security and privacy. These fake accounts are often behind impersonation, online scams, the spread of misinformation, and various other harmful activities.

Right now, most platforms rely on manual verification and user reporting to catch fake profiles. The problem is, this approach is slow, labour-intensive, and often struggles to keep up with increasingly sophisticated fake accounts. What's really needed is a smarter, automated system that can detect these profiles more accurately and efficiently.

That's exactly what this project sets out to do. "AI-Based Fake Profile Detection for Instagram Using Machine Learning" is a system designed to spot suspicious Instagram accounts using AI and machine learning techniques. It works by examining key features from a user's profile, pulling out the most relevant characteristics, and running them through a trained machine learning model that classifies the account as either genuine or fake.

The system is built using Python and the Flask framework, along with various machine learning libraries. It comes with a simple, easy-to-use interface where users can input profile details and instantly get a prediction. Beyond just labeling a profile as real or fake, the system also provides a confidence score, a risk level, and a brief explanation – all aimed at helping users better understand and trust the results.

By automating this process, the system significantly cuts down on the need for manual checks and offers a much faster way to flag potentially fake accounts. It's a practical example of how artificial intelligence can be applied to real-world cybersecurity and social media safety challenges. Looking ahead, there's plenty of room to build on this work – for instance, by incorporating deep learning for image analysis, natural language processing for text-based cues, real-time data analysis, and more advanced behavioural detection techniques.

INTRODUCTION

The rapid growth of internet technologies has transformed the way people communicate, share information, and interact with one another. Social media has become a core part of everyday life, giving users the ability to connect with people around the world, share personal experiences, and build their own digital identities. Among the many platforms available today, Instagram remains one of the most popular, allowing users to share photos, videos, stories, and various other forms of content.

However, as social media usage continues to grow, so do the security challenges that come with it. One of the biggest concerns is the rise of fake profiles — accounts created using false information, stolen identities, or automated bots. These fake accounts are frequently used for online fraud, impersonation, spreading misinformation, cyberbullying, and other privacy violations.

Detecting fake profiles has become increasingly difficult, mainly because of the sheer number of users on these platforms and the constantly evolving tactics used by those creating fake accounts. Traditional detection methods, which rely heavily on manual verification and user reports, tend to be slow and often fail to catch fake profiles accurately or in time.

This is where Artificial Intelligence (AI) and Machine Learning (ML) come in as a more effective solution. Machine Learning algorithms are capable of analysing various profile-related features, uncovering patterns that aren't easily visible to

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 7| July 2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

the human eye, and classifying accounts based on data the model has already learned from.

This project, titled "AI-Based Fake Profile Detection for Instagram Using Machine Learning," focuses on building an intelligent system that can identify suspicious Instagram profiles by analysing their available characteristics. The system uses Machine Learning techniques to classify profiles as either genuine or fake, and it presents its predictions along with a confidence score and a clear explanation to help users understand the results.

Overall, this project highlights how Artificial Intelligence can be practically applied in the fields of cybersecurity and social media protection, offering a faster and more automated way to identify suspicious online accounts

BACKGROUND OF THE STUDY

Social media has grown into one of the most powerful communication tools of our time, connecting millions of users across the globe. Platforms like Instagram give people the ability to create profiles, share content, interact with others, and become part of larger online communities.

But this same openness that makes social media so accessible also brings certain security risks. One of the most common problems is the creation of fake profiles, which can be set up easily using false or misleading information. In many cases, these fake accounts look almost identical to real ones, making it hard to tell them apart just by looking at them manually.

There are several reasons why fake Instagram profiles are created, including:

- Identity theft and impersonation
- Online scams and fraudulent activities
- Spreading false or misleading information
- Artificially boosting follower counts
- Promoting harmful or malicious content
- Collecting personal information from unsuspecting users

In the early days, social media platforms mostly depended on manual reporting and verification to deal with fake accounts. But with millions of new accounts being created every single day, this approach simply isn't enough anymore to keep up with the scale of the problem.

This is where Machine Learning offers a much more practical solution. Instead of relying on manual checks, it can automatically learn patterns directly from profile data. By examining factors such as username characteristics, how complete a profile is, and overall activity patterns, Machine Learning models can flag suspicious accounts far more efficiently than traditional methods.

With this in mind, the proposed system focuses on building an AI-based detection approach that helps users identify potentially fake Instagram profiles with greater speed and accuracy.

OBJECTIVE OF THE PROJECT

The main objectives of this project are as follows:

- To develop an AI-based system capable of detecting fake Instagram profiles.
- To analyse various profile-related features using Machine Learning techniques.
- To classify profiles into two categories — genuine and fake.
- To reduce the amount of manual effort currently required for profile verification.
- To provide prediction results along with clear confidence scores.
- To generate easy-to-understand explanations for how each detection decision is made.
- To ultimately help improve overall security and trust within social media platforms.

SCOPE AND SIGNIFICANCE

Scope

The scope of this project centres around developing a Machine Learning-based application that analyses Instagram profile features to detect potentially suspicious accounts.

More specifically, the system focuses on the following areas:

- Extracting meaningful features from available profile information
- Applying Machine Learning techniques to classify profiles
- Automatically generating predictions based on the trained model
- Presenting results in a clear, user-friendly manner

That said, there's plenty of room for this project to grow further down the line. Some possible future improvements include:

- Incorporating profile image analysis
- Using Natural Language Processing to analyse captions and comments
- Integrating real-time data through social media APIs
- Adding more advanced behavioural analysis techniques to strengthen detection accuracy

Significance

This project holds real significance because it offers an automated solution to a growing problem — the detection of fake profiles. Beyond just solving a technical challenge, it also demonstrates how Artificial Intelligence can be practically applied to real-world cybersecurity issues, ultimately helping users make safer, more informed decisions when interacting on social media.

DOMAIN COVERAGE

The AI-Based Fake Profile Detection System covers multiple domains where user identity verification and online security are important. These include:

- **Social Media Platforms** – Detects fake, bot, and impersonation accounts.
- **Cybersecurity** – Prevents phishing, identity theft, and online fraud.
- **Artificial Intelligence & Machine Learning** – Uses AI algorithms to classify genuine and fake profiles.
- **Data Mining & Analytics** – Extracts and analyses user behaviour and profile features.
- **Natural Language Processing (NLP)** – Examines profile descriptions, posts, and comments for suspicious content.

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 7| July 2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

- **E-commerce & Online Marketplaces** – Identifies fraudulent seller and buyer accounts.
- **Online Banking & Financial Services** – Enhances customer verification and fraud detection.
- **Law Enforcement & Digital Forensics** – Assists in investigating cybercrime and fake identities.

LITERATURE REVIEW

A literature survey plays a crucial role in any research project, as it offers a solid understanding of the work that's already been done in the field. It helps identify the different methods, technologies, strengths, and limitations found in previous research, which in turn allows for the development of a more refined and improved solution.

With platforms like Instagram, Facebook, Twitter (X), and LinkedIn growing at such a rapid pace, fake profiles have increasingly become a serious concern. These accounts are often created with malicious intent — ranging from identity theft and phishing to spamming, spreading misinformation, financial fraud, and cyberbullying. Given the massive number of users on these platforms, combined with the constantly evolving tactics used by attackers, manually detecting such profiles has become an incredibly difficult task.

Over the years, researchers have explored several different approaches to tackle this problem, including rule-based methods, Machine Learning algorithms, Deep Learning models, and broader AI-based techniques. Among all of these, Machine Learning has stood out as one of the most effective approaches, largely because of its ability to automatically learn patterns from data and use them to classify profiles with a good degree of accuracy.

This chapter reviews the existing research related to fake profile detection, discusses various Machine Learning algorithms, analyzes existing systems, identifies research gaps, and explains how the proposed system addresses those limitations.

METHODOLOGY

The proposed methodology follows a series of well-defined steps, outlined below:

Step 1: Data Collection

The process begins by gathering profile-related information from a prepared dataset, which includes examples of both genuine and fake Instagram profiles.

Step 2: Data Pre-processing

Once collected, the data is cleaned up — duplicate entries and missing values are removed — and then converted into a format that's suitable for Machine Learning.

Step 3: Feature Extraction

Next, important profile characteristics are extracted, such as:

- Username length
- Number of digits used
- Presence of special characters
- Overall username pattern
- Profile completeness

These extracted features are then converted into numerical values, making them ready for model training.

Step 4: Model Training

At this stage, the processed dataset is split into training and testing sets. A Machine Learning algorithm is then trained on this data, learning the underlying relationship between profile features and the corresponding profile categories.

Step 5: Prediction

Once the model is trained, it can be put to use — when a user enters profile information into the system, the model predicts whether that profile is genuine or fake.

Step 6: Result Generation

finally, the system presents the results in a clear and structured way, displaying:

- The prediction outcome
- A confidence score
- A risk level
- An explanation for the decision
- A relevant recommendation

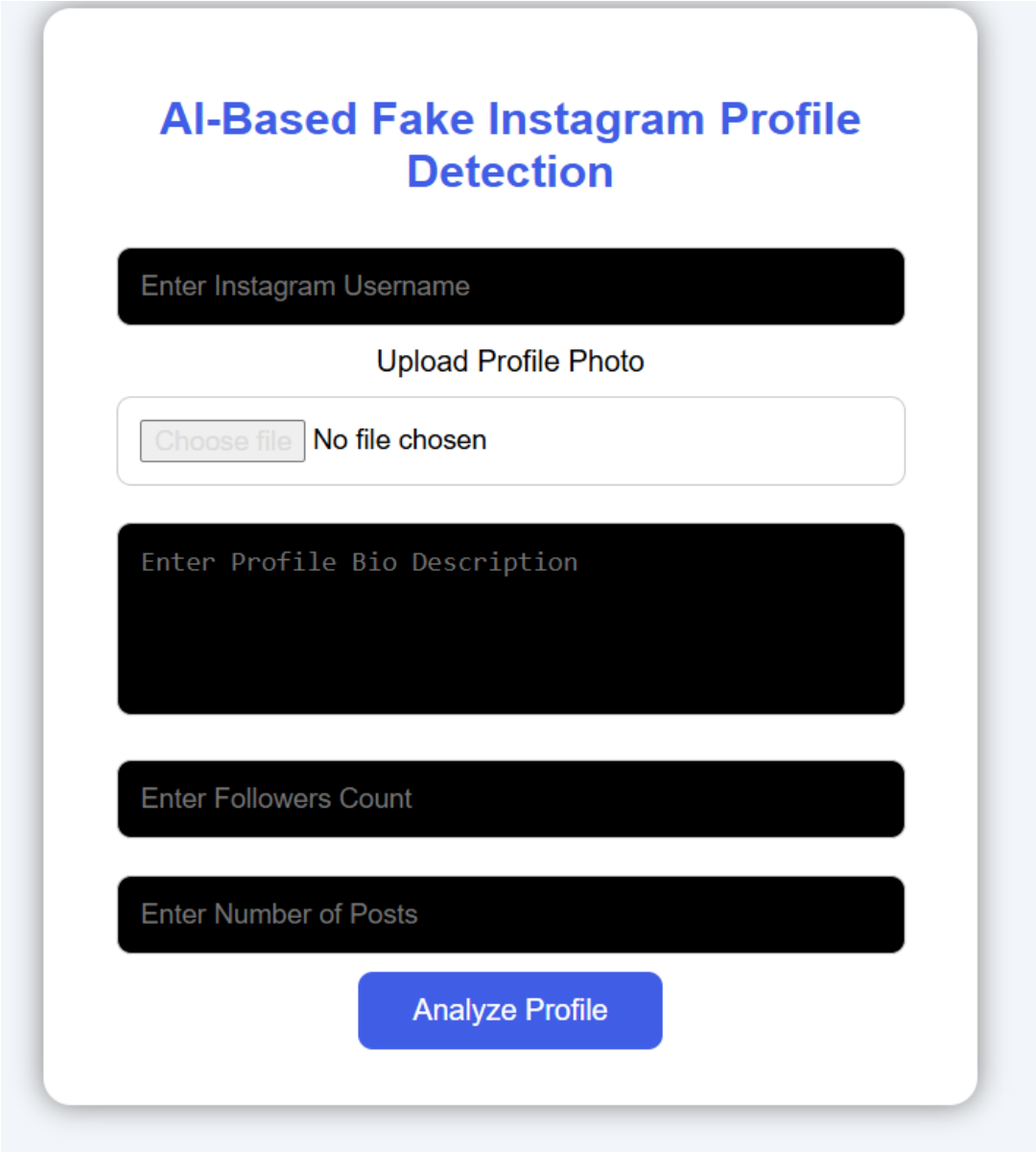
SYSTEM IMPLEMENTATION

Implementation is one of the most crucial phases in software development, as it's the stage where the system design actually comes to life and turns into a fully functional application. During this phase, the proposed AI-Based Fake Profile Detection for Instagram Using Machine Learning system is built using suitable programming languages, frameworks, and Machine Learning libraries.

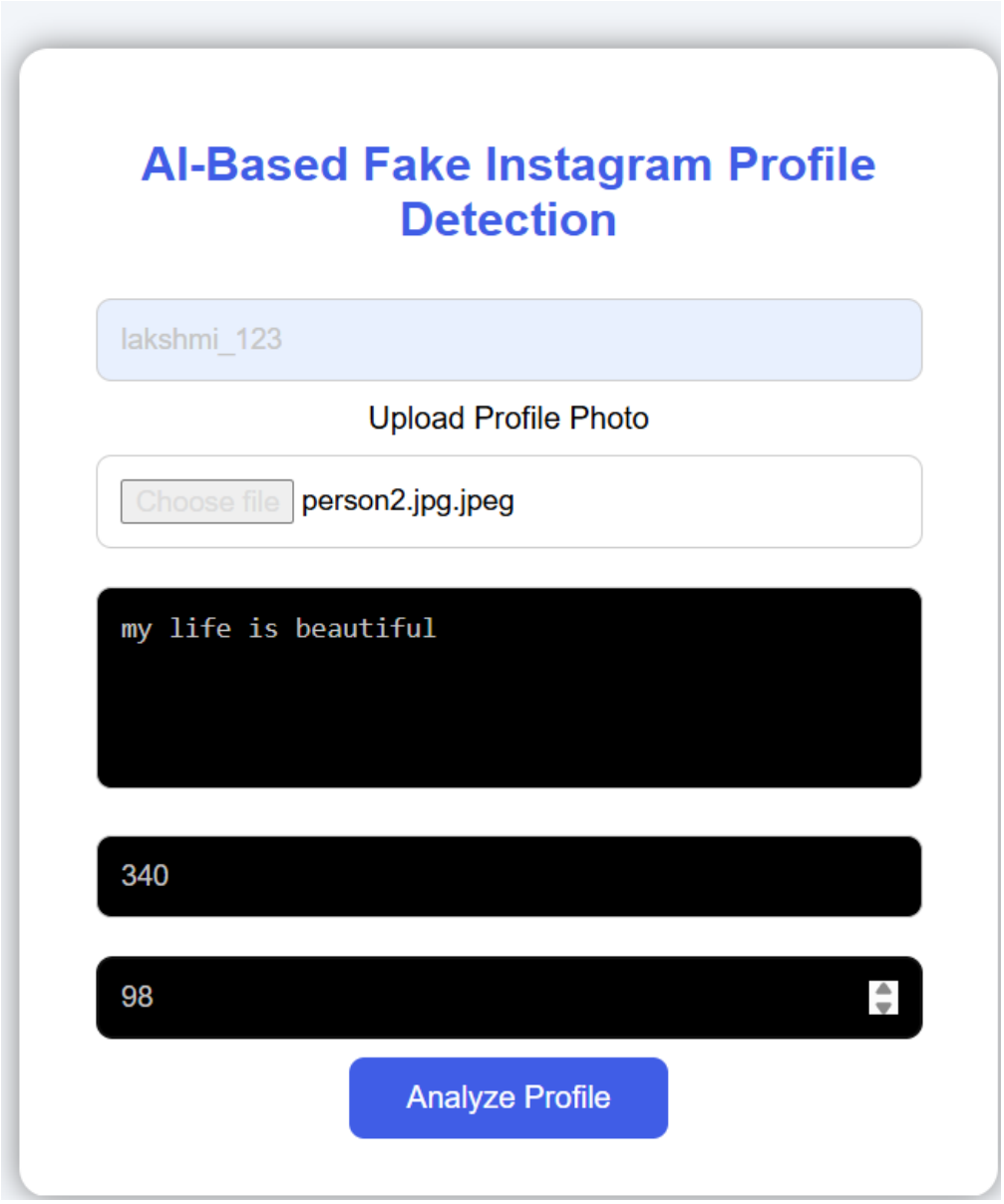
The implementation process covers several important steps, including dataset preparation, feature extraction, model training, integrating the Machine Learning model with a web application, and developing the user interface. Throughout this process, the main focus remains on creating a system that's genuinely easy for users to work with — allowing them to detect fake Instagram profiles simply by analyzing key profile-related features.

The application itself is built using Python and Flask, along with HTML and CSS for the frontend, while Machine Learning libraries such as Scikit-learn, Pandas, and NumPy handle the core data processing and model-building tasks. Once trained, the model is integrated directly into the Flask application, allowing it to generate real-time prediction results based on whatever input the user provides.

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 7| July 2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188



ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 7| July 2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188



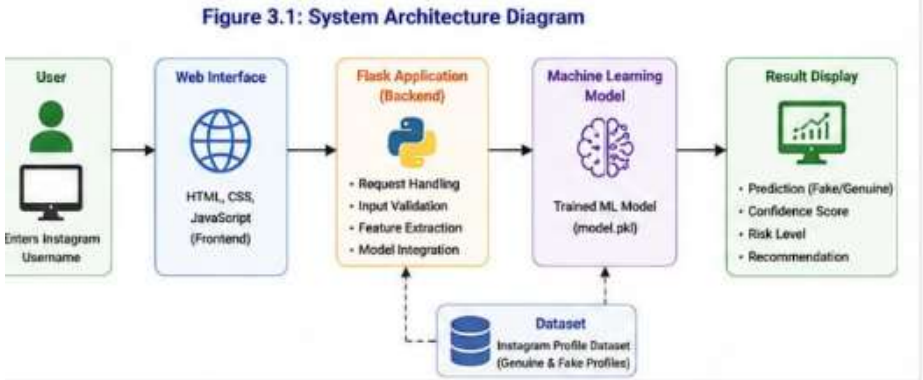
SYSTEM ARCHITECTURE

The overall architecture of the proposed system is built around the following key components:

- User Interface
- Input Validation
- Feature Extraction Module
- Machine Learning Model
- Prediction Engine
- Result Display Module

Together, these components are designed to work seamlessly with one another, ensuring smooth and efficient interaction between the user and the underlying Machine Learning model.

Figure 3.1: System Architecture Diagram



ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 7| July 2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

Tools and Technologies Used

Bringing the proposed system to life required a combination of different software tools and technologies, each playing its own role in areas like data processing, model training, web development, and result visualization.

Python

Python served as the primary programming language for this project. Its simple, readable syntax combined with a rich set of libraries makes it a natural choice for Artificial Intelligence and Machine Learning applications.

Flask Framework

Flask, a lightweight Python web framework, was used to build the web application itself. It takes care of handling user requests, communicating with the Machine Learning model behind the scenes, and displaying the final prediction results to the user.

HTML

HTML (Hypertext Mark-up Language) was used to structure the web pages, including both the home page and the results page.

CSS

CSS (Cascading Style Sheets) was brought in to enhance the overall look and feel of the application, giving it a clean, attractive, and responsive interface.

JavaScript

JavaScript was used to handle client-side interactions, making the web pages feel more dynamic and responsive to user actions.

Scikit-learn

Scikit-learn, one of the most popular Machine Learning libraries, was used for training, testing, and evaluating the classification model at the core of this system.

Pandas

Pandas played a key role during the data pre-processing stage, helping with reading, organizing, and manipulating the dataset efficiently.

NumPy

NumPy provided the numerical computation capabilities needed throughout the data processing and Machine Learning workflow.

Joblib

Joblib was used to save and later load the trained Machine Learning model (model.pkl), making it easy to reuse the model for generating predictions.

Visual Studio Code

Finally, Visual Studio Code served as the Integrated Development Environment (IDE) throughout the project, providing a smooth environment for writing, testing, and debugging the code.

Table 4.2: Tools and Technologies

Tool/Technology	Purpose
Python	Programming Language
Flask	Web Framework
HTML	Web Page Structure
CSS	User Interface Design
JavaScript	Client-side Interaction
Scikit-learn	Machine Learning Model
Pandas	Data Processing
NumPy	Numerical Computation
Joblib	Model Saving and Loading
Visual Studio Code	Development Environment

ADVANTAGES

Here's a breakdown of the key advantages of an AI/ML-based fake profile detection system for Instagram:

1. Scalability

- Instagram has over 2 billion users generating millions of new accounts daily — manual review is impossible at this scale.
- ML models can screen accounts in real time, processing volumes no human team could handle.

2. Speed and Real-Time Detection

- Fake profiles can be flagged at the moment of creation or shortly after, based on signup patterns, before they cause harm (spam, scams, follower fraud).
- Faster response reduces the window for bad actors to exploit the platform.

3. Pattern Recognition Beyond Human Capability

- ML models can detect subtle, non-obvious correlations across dozens of features simultaneously (posting frequency, follower/following ratio, profile completeness, engagement patterns, device/IP metadata) that a human reviewer would miss.
- Can identify coordinated bot networks based on shared behavioral fingerprints, not just individual account traits.

4. Continuous Learning and Adaptability

- Models can be retrained as fake-profile tactics evolve (adversarial adaptation), unlike static rule-based filters that quickly become outdated.
- Techniques like online learning let the system adjust to new fraud patterns without a full system overhaul.

5. Reduced Manual Moderation Burden

- Automates the bulk of first-pass screening, freeing human moderators to focus on ambiguous or appealed cases.
- Lowers operational costs compared to large-scale manual review teams.

6. Multi-Modal Analysis

- Can combine text (bio analysis, NLP on captions/comments), image analysis (reverse image search, GAN-generated face detection, profile picture authenticity), and behavioral/network data (graph-based analysis of follower networks) for more robust detection than any single signal.

7. Improved Accuracy Over Time

- Supervised learning models improve as more labeled data (confirmed fake vs. real accounts) becomes available, reducing false positives/negatives.

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 7| July 2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

- Ensemble methods (combining multiple classifiers) can boost precision.

8. Proactive Fraud and Scam Prevention

- Helps curb catfishing, romance scams, phishing links, fake engagement (bought followers/likes), and impersonation before users are victimized.
- Protects brand advertisers from wasted ad spend on bot-inflated audiences.

Limitations

While the developed system performs well overall, it's important to acknowledge a few limitations that came up along the way:

- The system currently analyses only a limited set of profile features
- The model was trained on a custom dataset rather than real, live Instagram data
- Although profile images can be uploaded, they aren't actually analysed using image processing techniques
- The application doesn't automatically access live Instagram profiles on its own
- Prediction accuracy is closely tied to the quality and diversity of the training dataset used
- Factors like verified badges, account age, comments, and engagement rate aren't currently taken into account
- The system, as it stands, only supports profile-level analysis rather than a broader, more comprehensive evaluation

Future Enhancement

That said, there's plenty of room to build on what's already been developed. Some possible future enhancements include:

- Integrating the Instagram Graph API (where permitted) to enable real-time profile analysis
- Expanding the dataset to improve overall prediction accuracy
- Applying Deep Learning models, such as CNNs or Transformers, for more advanced detection capabilities
- Incorporating Computer Vision techniques to analyse profile images directly
- Adding more profile features into the mix, such as following count, engagement rate, hashtags, and account age
- Developing a mobile version of the application using Flutter or Android Studio
- Supporting multilingual profile analysis for broader usability
- Introducing an administrator dashboard to help monitor suspicious accounts more effectively
- Deploying the application on cloud platforms like AWS, Azure, or Google Cloud for wider accessibility
- Strengthening the explainable AI features by offering even more detailed reasoning behind each prediction

Conclusion

The AI-Based Fake Profile Detection for Instagram Using Machine Learning project was successfully designed, developed, and implemented to identify whether an Instagram profile is genuine or fake. The system utilizes a Random Forest Machine Learning algorithm to analyze various profile attributes such as username characteristics, bio description, follower count, and number of posts. Based on these extracted features, the trained model predicts the authenticity of a profile.

A user-friendly web application was developed using the Flask framework, allowing users to enter profile details and instantly receive the prediction result. The system also displays the confidence score, risk level, and AI-generated explanation, making the output easy to understand.

The trained model achieved high accuracy on the prepared dataset and demonstrated the capability to classify genuine and fake profiles effectively. The project highlights how Machine Learning techniques can assist in improving online security by identifying suspicious social media accounts. The objectives of the project were successfully achieved, and the developed system can serve as a foundation for future intelligent fake profile detection applications.

REFERENCES

The following resources were referred to throughout the project to build a solid understanding of Machine Learning concepts, Python programming, Flask web development, and fake profile detection techniques.

Books

1. *Ian Good fellow, Yoshua Bengio, and Aaron Courville, Deep Learning, MIT Press, 2016.*
2. *Aurélien Géron, Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow, 3rd Edition, O'Reilly Media, 2022.*
3. *Ethem Alpaydin, Introduction to Machine Learning, 4th Edition, MIT Press, 2020.*
4. *Tom Mitchell, Machine Learning, McGraw-Hill Education.*
5. *Jake VanderPlas, Python Data Science Handbook, O'Reilly Media.*

Research Papers

1. *Detecting Fake Accounts on Social Networks Using Machine Learning Techniques.*
2. *Machine Learning Approaches for Fake Profile Detection in Social Media.*
3. *Social Media Fake Account Detection Using Artificial Intelligence.*
4. *Fake Profile Identification Using Random Forest Classification.*
5. *AI-Based User Behaviour Analysis for Social Network Security.*

Websites

- *https://www.python.org*
- *https://flask.palletsprojects.com*
- *https://scikit-learn.org*
- *https://pandas.pydata.org*
- *https://numpy.org*
- *https://matplotlib.org*
- *https://github.com*
- *https://code.visualstudio.com*