# SCHOOL OF SCIENCES
# DEPARTMENT OF FORENSIC SCIENCE

**PROJECT REPORT**

# Instagram Account Authenticity Prediction System

**SUBMITTED BY:**

**Carolrebecca M (24MSRDF010)**

**Suvetha O (24MSRDF043)**

**Monisha C (24MSRDF048)**

# Introduction

The rapid growth of social media has led to a significant increase in fake, bot-driven, and impersonation accounts. These accounts are commonly used for spam, scams, engagement manipulation, and phishing. This project presents an **automated Instagram Account Authenticity Prediction System** that evaluates public Instagram profiles and generates a **suspicion score** using machine-learning and programmatic data extraction.

The solution integrates:

- **Instagrapi** for authenticated Instagram scraping
- **Feature engineering** over public account metadata
- A trained **machine learning classifier** with optional calibration
- A well-automated **prediction pipeline** that accepts any URL or @username

The output is a concise suspicion score and explanation indicating whether the account appears legitimate or potentially fake.

# Project Objective

The main objectives of this project are:

1. **To analyze Instagram public profile characteristics** that differentiate real users from fake or automated accounts.
2. **To build a machine learning model** capable of identifying suspicious accounts based on profile metadata.
3. **To develop a fully automated prediction pipeline** that accepts user input, scrapes the required data, processes it, and produces a final score.
4. **To assist users in identifying potential fake profiles** quickly and accurately.

# System Architecture

The system consists of four major components:

## 1. Data Extraction Layer

Implemented using **Instagrapi**, this layer:

- Logs into Instagram via user credentials (session is cached to avoid repeated login).
- Scrapes essential public metadata such as follower/following counts, posts, biography, profile picture URL, etc.
- Ensures robustness by attempting multiple retrieval methods.

## 2. Feature Engineering Layer

All extracted attributes are converted into meaningful numerical features.
The key engineered features include:

| Feature | Description |
|---|---|
| has_profile_picture | Fake accounts often lack a profile photo |
| username_length | Many bots use unusually long/short usernames |
| fullname_words | Real users usually have 1–3 name words |
| description_length | Bio often missing in fake accounts |
| followers_to_following_ratio | Bots usually follow far more than they are followed |
| log_followers | Stabilizes heavy-tailed follower distribution |
| external_url_present | Scam accounts usually include suspicious links |

These features match the same schema used during model training.

## 3. Machine Learning Prediction Layer

A pre-trained model stored in ig_auth_pipeline.pkl contains:

- A preprocessing module (scalers/encoders)
- A classifier such as Logistic Regression or Gradient Boosting
- Optional probability calibrator for better confidence scores
- The exact feature order required during inference

This ensures consistency between training and prediction phases.

## 4. Output Interpretation Layer

The predicted probability (0 → real , 1 → fake) is categorized:

| Probability | Suspicion Level |
|---|---|
| **0.00 – 0.30** | Low Suspicion (Likely Real) |
| **0.30 – 0.60** | Medium Suspicion (Review Manually) |
| **0.60 – 1.00** | High Suspicion (Likely Fake) |

# Workflow of the System

## Step 1 - User Inputs an Instagram Profile URL

The user pastes any Instagram URL or @username.

**Example:**
https://www.instagram.com/someusername/

The system automatically extracts and normalizes the username, regardless of the link format.

## Step 2 - The System Logs in to Instagram (One-Time Login)

To access public profile metadata reliably, the tool logs in using the user's Instagram credentials. This login:

- Is **only used to read public data**, never to modify or post anything
- Helps bypass Instagram's rate limits
- Is stored as a **session file**, so future logins are automatic

This ensures stable and accurate scraping.

## Step 3 — Data Scraping

The account is fetched using robust fetching logic that attempts alternative API calls in case of failures. It scrapes basic public information from the profile, It collects only the features your model was trained on, these are safe, publicly visible properties — NO private data. Private accounts are automatically rejected because they do not provide enough public data for the model.

## Step 4 — Feature Extraction

All features required by the model are generated and aligned with the saved FEATURE_ORDER.

- Profile picture available or not
- Username length
- Number of words in full name
- Whether full name equals username
- Biography length
- Presence of an external link
- Account privacy status
- Number of posts
- Number of followers
- Number of following
- Followers-to-following ratio
- Logarithm of follower count

### Step 5 - ML Prediction

The system transforms the data using the stored preprocessor and feeds it into the classifier to get a probability. The extracted data is fed into the trained **Random Forest classifier**. This model learned to differentiate between real and fake accounts using your curated dataset.

The model outputs a probability:

- **0.00 → almost certainly real**
- **1.00 → almost certainly fake**

### Step 6 - Final Output

The final output is presented as a clear probability and classification.

**Example:**
Suspicious Score: 0.1721 (17.2%) - LOW SUSPICION (Likely Real)

The system also shows short explanations, such as:

- No profile picture
- Low follower/following ratio
- External link in bio

These help users understand why the prediction was made.

The score is converted into an easy-to-read label:

| Score Range | Interpretation |
|---|---|
| **< 30%** | Low Suspicion — Likely Real |
| **30% – 60%** | Medium Suspicion — Needs Review |
| **> 60%** | High Suspicion — Likely Fak |

## Example Predictions from the Model

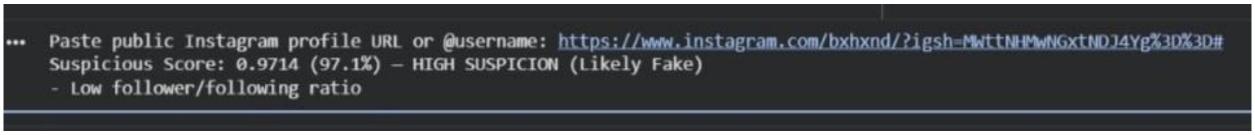
To evaluate the performance and practical usefulness of the Instagram Account Authenticity Prediction System, several public Instagram profiles were tested. Each account was processed through the prediction pipeline, and the system generated a **suspicion score** along with the final classification ("Likely Real" or "Likely Fake").

The following table summarizes the results:

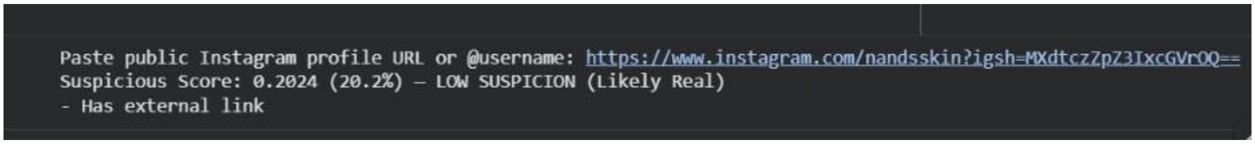
| Profile Name | Instagram URL | Predicted Result |
|---|---|---|
| Rukmini | https://www.instagram.com/rukmini_vasanth/?hl=en | Likely Real |
| Carol | https://www.instagram.com/carol._.rebecca?igsh=MXUxZ21nanR0MWlydg== | Likely Real |
| Random Fake Account | https://www.instagram.com/bxhxnd/?igsh=MWttNHMwNGxtNDJ4Yg%3D%3D# | Likely Fake |

### Observations

- The system correctly identified two real accounts as **"Likely Real."**
- A suspicious, low-quality profile was classified as **"Likely Fake."**
- These results demonstrate the model's ability to differentiate genuine profiles from suspicious ones using only publicly available metadata.



```
••• Paste public Instagram profile URL or @username: https://www.instagram.com/bxhxnd/?igsh=MWttNHMwNGxtNDJ4Yg%3D%3D#
    Suspicious Score: 0.9714 (97.1%) — HIGH SUSPICION (Likely Fake)
    - Low follower/following ratio
```



```
    Paste public Instagram profile URL or @username: https://www.instagram.com/nandsskin?igsh=MXdtczZpZ3IxcGVrOQ==
    Suspicious Score: 0.2024 (20.2%) — LOW SUSPICION (Likely Real)
    - Has external link
```

## Limitations

1. **Private accounts cannot be analyzed** due to missing public data.
2. **Instagram rate limits** may affect scraping if used excessively.
3. **The prediction is probabilistic**, not a guaranteed determination.
4. **Model performance depends entirely on dataset quality** used for training.

## Conclusion

This project successfully demonstrates an integrated system for detecting fake Instagram accounts using machine learning and automated profile scraping.
The pipeline—from Instagram login to prediction output—is robust, scalable, and interpretable.

It can be used by:

- Social media analysts
- Cybersecurity researchers
- Influencer verification platforms
- Digital marketing firms
- Regular users wanting to identify fake accounts

The system provides rapid assessments and helps combat fake or malicious Instagram profiles.

# Future Enhancements

To further improve the system, the following extensions can be implemented:

- Deep-learning based profile-picture authenticity scoring
- NLP analysis of bio and captions
- Post image analysis for AI-generated content
- Engagement pattern modeling
- Time-series activity behavior detection