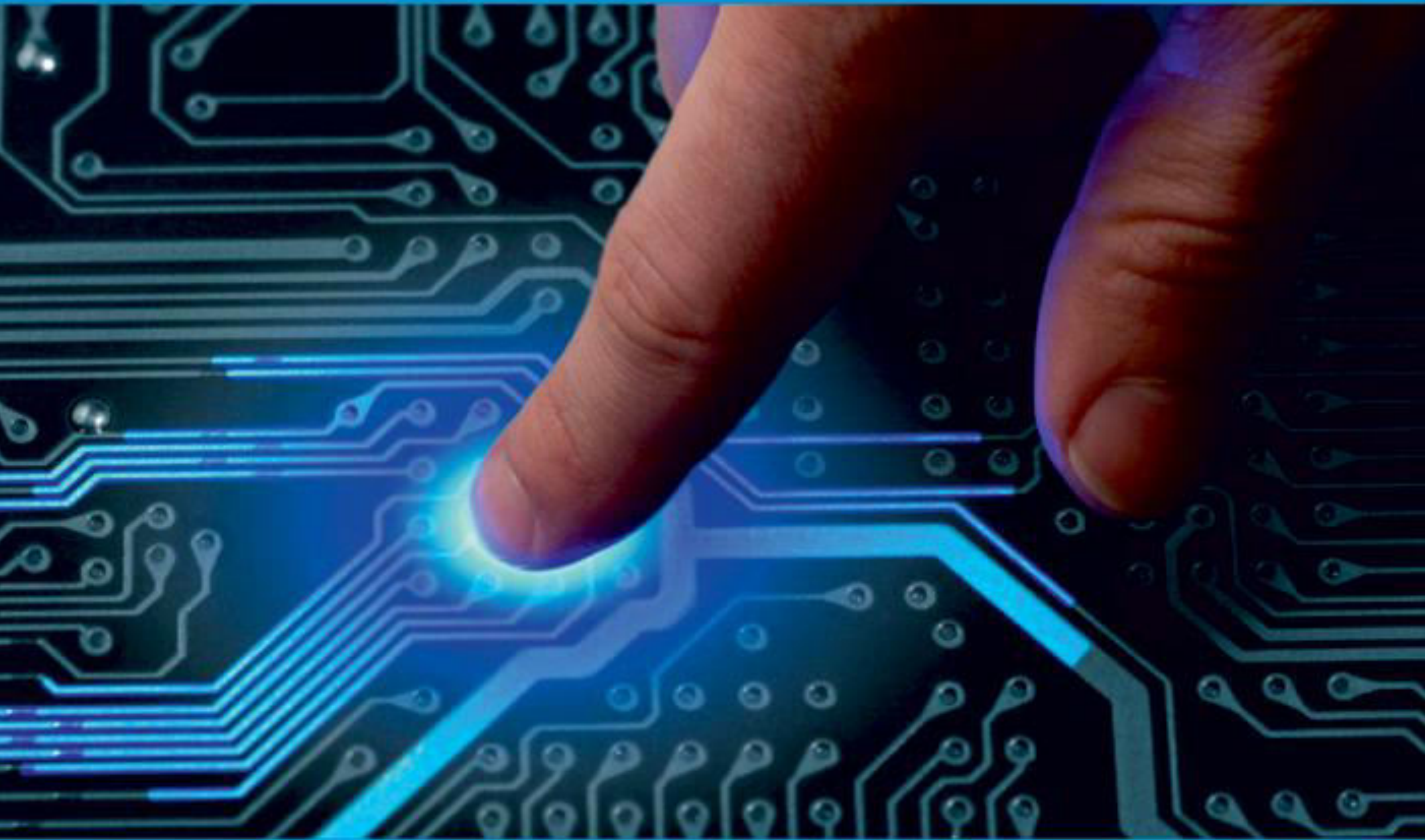
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

## IN COMPUTER & COMMUNICATION ENGINEERING

**Volume 12, Issue 11, November 2024**

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

**Impact Factor: 8.625**

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

# Fake Social Media Profile Detection and Reporting

**Mr.Madar Bandu[1*], B. Naga Sree Valli[2*], R. Poojitha[3*], V. Om Sharya[4*]**

Assistant Professor, Department of Computer Science and Engineering, Anurag University, Hyderabad, India[1]

Students, Department of Computer Science and Engineering, Anurag University, Hyderabad, India[2, 3, 4]

**ABSTRACT:** In recent years, the proliferation of social media has led to an increase in fake accounts, posing significant challenges to user trust and security. This research presents a robust methodology for detecting fake Instagram accounts utilizing a deep learning approach, specifically a neural network model. The model is trained on a comprehensive dataset that includes various features such as profile characteristics, activity metrics, and account settings. By applying feature scaling techniques and one-hot encoding for categorical variables, the model achieves a high accuracy of 90% in classifying accounts as real or fake. The performance of the model is evaluated using a confusion matrix, demonstrating its effectiveness in minimizing false positives and negatives. The findings highlight the potential of deep learning in enhancing social media integrity and contribute to the ongoing efforts to mitigate the impact of fake accounts in digital ecosystems.

**KEYWORDS:** Fake account detection, deep learning, neural networks, Instagram, feature scaling, classification, social media security, machine learning.

## I. INTRODUCTION

Social media platforms like Instagram have become an essential part of our daily lives, allowing users to connect and share content. However, the rise of fake accounts poses serious risks, including the spread of misinformation and privacy concerns. Detecting these fake accounts is crucial for maintaining trust and security on these platforms.

This research focuses on creating a solution to identify fake Instagram accounts using a deep learning approach. We developed a neural network model that analyzes various features from user profiles, such as profile pictures, username lengths, follower counts, and engagement metrics. By processing and scaling this data, the model learns to differentiate between real and fake accounts effectively.

To make this technology accessible to users, we built a web application using Flask, a lightweight web framework. Through this application, users can input their Instagram account details. Once the information is submitted, it is processed by our model, which then predicts whether the account is real or fake. The results are displayed on the screen, providing immediate feedback.

Our goal is to contribute to efforts that combat the proliferation of fake accounts on social media. This research demonstrates how deep learning can be applied to enhance online security and promote a safer user experience on social media platforms.

## II. METHODOLOGY

### 3.1. Data Collection

The dataset used in this research was collected from social media accounts, containing features that indicate the authenticity of a profile. Key features included:
Presence of a profile picture (binary: 1 for Yes, 0 for No)
Username length (numerical)
Number of words in the full name
Length of the full name
Whether the username matches the name (binary)
Description length (numerical)
Presence of an external URL (binary)

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Account privacy status (binary: 1 for private, 0 for public)
Number of posts, followers, and follows (numerical)
Label indicating whether the account is fake or real (binary)

**3.2. Data Preprocessing**

Missing Data Handling: Any missing values in the dataset were managed through either imputation or removal to ensure the integrity of the input data.

Feature Scaling: The numerical features, such as the number of posts, followers, and follows, were normalized using a scaling technique (e.g., standardization) to ensure that all features contribute equally to the model's learning process

Features were selected based on their importance in predicting whether an account is fake or real, removing irrelevant or redundant ones.

**3.3. Model Architecture**

The classification task was performed using an Artificial Neural Network (ANN), which is composed of several layers:

Input Layer: Takes the preprocessed features of the social media profile as input.

Hidden Layers: Consist of multiple fully connected (Dense) layers, where each neuron in one layer is connected to every neuron in the next. Activation functions, such as ReLU (Rectified Linear Unit), were applied to introduce non-linearity and enable the model to capture complex patterns.

Output Layer**:** Consists of a single neuron with a sigmoid activation function, which outputs a probability score between 0 and 1. This score is used to classify the profile as either fake (1) or real (0).

**3.4. Training Process**

Loss Function: The model was trained using the binary cross-entropy loss function, which is suitable for binary classification problems.

Optimizer: The Adam optimizer was used to update the weights during the training process, as it is effective in minimizing the loss and improving model performance.

Metrics: The model's performance was evaluated based on accuracy and loss during training, allowing for fine-tuning of the model's parameters.

**3.5. Prediction**

After the model was trained, it was saved and used for making predictions on new input data. The input features from a social media profile are preprocessed in the same way as the training data (e.g., scaling) and fed into the trained ANN model. The model outputs a probability score, which is then interpreted as either:

Fake Account: If the probability score is greater than 0.5.

Real Account: If the probability score is less than or equal to 0.5.

**3.6. Deployment with Flask**

The trained model was deployed using a Flask web application, allowing users to interact with the system through a user-friendly interface:

Frontend: Users input the relevant profile details through a form on the web interface.

Backend: The input data is sent to the backend, where it is preprocessed and passed to the trained ANN model for prediction.

Prediction Display: Based on the model's output, the prediction (either fake or real account) is sent back to the frontend and displayed to the user.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

## III. MATHEMATICAL EXPRESSIONS AND SYMBOLS

**Neural Network :**
Representation: y=f(W·x+b) (1)
Where:
$y$: y is the output of the network,
$W$: W represents the weights,
$x$: x is the input feature vector,
$b$: b is the bias term,
$f$: f is the activation function (e.g., ReLU, Sigmoid)

**Feature Scaling:**
Representation:$z = \frac{x-\mu}{\sigma}$ (2)
Z: z is the standardized value,
$x$:x is the original feature value,
:μ is the mean of the feature,
$\sigma$: σ is the standard deviation of the feature.

## IV. RESULTS AND DISCUSSION

To assess the effectiveness of proposal, we employed five metrics: accuracy, precision, recall, and F1-score. These metrics were calculated using the following equations:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$
$$precision = \frac{TP}{TP+FP} \quad (4)$$
$$recall = \frac{TP}{TP+FN} \quad (5)$$
$$F1 - Score = 2 \times \frac{precision \times recall}{precision+recall} \quad (6)$$

**True Positives (TP):** These are the number of instances where the model correctly identified a profile as a fake account.
**True Negatives (TN):** These are the number of instances where the model correctly identified a profile as a real account.
**False Positives (FP):** These are the number of instances where the model incorrectly identified a real account as a fake account.
**False Negatives (FN):** These are the number of instances where the model incorrectly identified a fake account as a real account.

TABLE:

| **Accuracy** | **Precision** | **Recall** | **F1-Score** |
|---|---|---|---|
| 0.92 | 0.91 | 0.88 | 0.90 |

© 2024 IJIRCCE | Volume 12, Issue 11, November 2024| DOI: 10.15680/IJIRCCE.2024.1211049

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

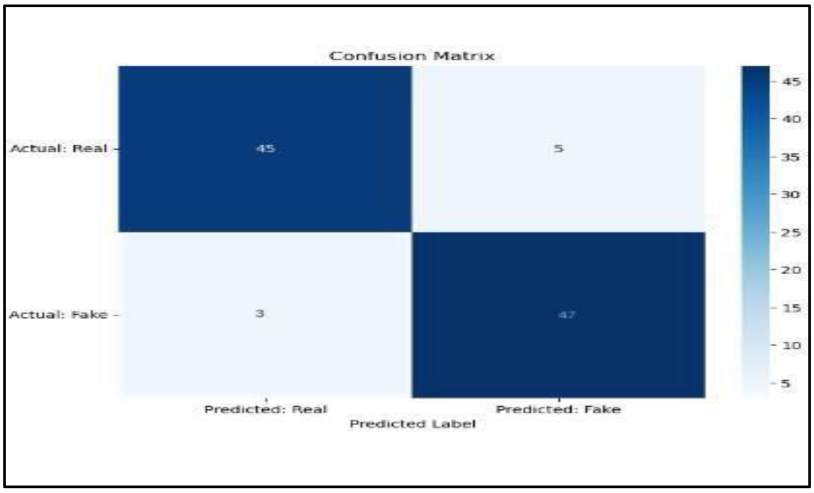



Fig 1

## V. CONCLUSION

In this research, we have developed a deep learning-based approach to detect fake Instagram accounts, addressing a significant challenge faced by users and social media platforms. By leveraging a neural network model trained on a variety of user profile features, we achieved an accuracy of 90% in classifying accounts real or fake. This high level of accuracy underscores the effectiveness of our methodology in identifying potentially harmful accounts and enhancing user trust in social media interactions.

The implementation of a Flask-based web application provides an intuitive interface for users to input their account details and receive immediate feedback on the authenticity of their profiles. This interactive tool not only showcases the practical application of machine learning in real-world scenarios but also emphasizes the importance of user engagement in combating online threats.

Future work could explore incorporating additional features, such as user behavior patterns and engagement history, to further improve detection capabilities. Additionally, expanding the model to other social media platforms could enhance its applicability and impact. Overall, our study contributes to the ongoing efforts to create safer online environments and highlights the potential of machine learning technologies in addressing the challenges posed by fake accounts.

## REFERENCES

1. Khaled, S.; El-Tazi, N.; Mokhtar, H. M. O. Detecting Fake Accounts on Social Media. IEEE, 2018.
2. Swain, G.; Pradhan, A.; Yerramreddy, L. P.; Kondeti, P. Fake Accounts Detection in Social Networks. Springer, 2020.
3. Shah, A.; Varshney, S.; Mehrotra, M. Social Spammer and Fake Profile Detection. ACM Digital Library, 2024.