iJRASET
International Journal For Research in Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 13 **Issue:** XII **Month of publication:** December 2025

**DOI:** https://doi.org/10.22214/ijraset.2025.76717

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 13 Issue XII Dec 2025- Available at www.ijraset.com*

# A Machine Learning Approach for Identifying Fake Accounts on Instagram

Prof. S.N.Pawar[1], Om Sudhakar Chaudhari[2], Khushal Sunil Lohar[3], Kundan Gajanan Patil[4], Prathamesh Sanjay Khairnar[5]

[1]*Project Guide, R. C. Patel Institute of Technology, DBATU, Shirpur, India*
[2, 3, 4, 5]*UG Students, R.C.Patel Institute of Technology, DBATU, Shirpur, India*

***Abstract:*** ***Instagram is an increasingly popular social media platform in the digital social media ecosystem; however, the issue of fake and robotic accounts gaining momentum is dangerous and accordingly presents misinformation, scams, and identity abuse as significant threats. This study introduces an Instagram Fake Profile Detection system built on machine learning to detect suspicious users accounts with the help of behavioral, numerical, and profile-based features. The given system derives significant predictors, such as the ratio of followers to followed, the articles of the biography, the number of media, the structuring of the usernames, the presence of the profile picture, and privacy settings, and learns the trends typical of fake profiles. Several machine learning models are trained and compared, such as the Random Forest, SVM, ANN, GRU, LSTM, and Hybrid Deep Learning architecture, to increase the prediction accuracy. It adopts a Flask-based web interface enabling real-time classification and visualization of the model outputs of any instagram user input. The findings demonstrate that the reliability of fake account detection is enhanced due to the integration of traditional ML and deep learning networks. In general, this work would be a valuable addition to the body of research because it presents a functional, automatized, and scalable method of enhancing security and trust within the user space of Instagram.***

***Keywords: Instagram Fake Account Detection, Machine Learning, Deep Learning, Hybrid Model, Social Media Security, Fake Profile Classification***

## I. INTRODUCTION

The social media represent a vital aspect of everyday communication, and Instagram is one of the most widely used platforms where people can share photos, communicate with others, and conduct business. But with the rising popularity there has been a massive rise in the number of fake and automated profiles. These accounts may be involved in bad practices like misinformation sharing, following in large numbers, pushing scams, and impersonation of real users. This conduct does not only decrease the authenticity of the platform but also poses risks to people and organizations. To overcome these, there has been an increasing demand of smart systems that will automatically detect counterfeit profiles with high precision. Conventional manual tools like reviewing profile pictures, usernames, or number of followers are slow, unreliable, and inefficient when it comes to massive fake accounts. Machine learning is a more effective approach, as it holds patterns of real user behavior and fake user behavior and uses these patterns to classify accounts. The current study introduces an Insta- Based Fake Profile Detection system using machine learning that can effectively extract a variety of features about a user such as follower and following ratio, length of bio, number of media, nature of user name, private/public setting and availability of profile picture. Classical machine learning models on the system include Random Forest, SVM, and ANN, as well as modern Deep Learning models, including GRU, LSTM, and a Hybrid architecture. To differentiate fake and authentic accounts, such models are trained on extracted numerical features.

A Web application on Flask is also designed to apply real-time prediction, where the end user can provide Instagram profile details and be returned with classification results instantly. The system also graphs the predictions made by each algorithm and allows a direct comparison of the model performance. The proposed concept, by integrating effective preprocessing, more than one ML/DL models, and a straightforward user interface is a scalable and practical solution to fake profiles on Instagram.

## II. OBJECTIVES

*1)* Create an automated system that can check fake Instagram accounts based on machine learning models without the need to manually check them.

*2)* To study major Instagram user features including follower and followings, posts, and biography, username pattern, and profile pictures to detect suspicious nature.

*3)* To develop a safe and precise prediction model with various algorithms such as the Random Forest, SVM, ANN, GRU, LSTM, and Hybrid Deep Learning.
*4)* To develop a user friendly web based application in which users can input profile information and immediately verify whether an instagram account is genuine or not.
*5)* To deliver easy visual output and model comparison charts that will assist users in comprehending predictions brought about by dissimilar algorithms.
*6)* Online safety: to inhibit the proliferation of social media fake accounts, scams, and misinformation.

## III. LITERATURE REVIEW

Detection of bogus social media accounts has been a research topic that has been actively pursued due to the fact that fake profiles and bots do not only tamper with the integrity of a platform, but also manipulate and propagate misinformation. Initial research engaged profile and behavioral features that are not automatic (follower/following ratio, posting frequency, bio content, username pattern, profile picture appearance) and used standard classifiers, including Random Forests, SVMs and logistic regression, to differentiate between authentic and fake users. The new Instagram-specific works verify that metadata indicators, such as ratio between followers and followed, bio length, and similarity between the usernames, are always in the category of the most significant predictors of fake accounts (Chelas, Routis and Roussaki, 2024). [1]

Studies of other platforms (e.g., Twitter) have indicated that a combination of profile-level features with temporal and content features has a higher degree of protection against manipulation (e.g., synthetic follower boosting or intermittent posting). They note defining that the behavioral cues (rate of posting, engagement velocity) record evidence on the activity of users that simple snapshots of the profile miss, and that are frequently more effective to detect when combined with profile metadata. Although not necessarily an Instagram-specific phenomena, these inter-platform studies are relevant to the problem of Instagram detection in more recent literature.

In the recent past, there is an increased use of hybrid and deep-learning solutions. As an example, a new architecture with the help of a Long Short-Term Memory (LSTM) network was suggested to categorize fake and real Instagram accounts and demonstrated a high accuracy on the real account. -world datasets (MDPI, 2024). [2] Hybrid architectures To learn sequential and static textual or behavioral patterns, hybrid architectures based on one or more long-term memory in LSTM with the efficiency of GRU, or on feature embedding and recurrent layers, are becoming more common. These architectures can be seen to excel in a bot-detection task over single-model baselines because they simultaneously exhibit the short- and long-high dependencies in user activity.

Explicitly named studies on Instagram detection indicate that ensemble and optimization-based pipelines (e.g. via metaheuristic feature selection plus a combination of several classifiers), are more effective in cross dataset accuracy and generalization. As an example, binary Grey Wolf Optimization (BGWO) and Particle Swarm Optimization (PSO) have been employed in feature selection and passed to classifiers, such as ANN, SVM, KNN and Logistic Regression resulting in the very high appearance of fake-account detection (Algorithms, 2024). [3]. Practical issues are also brought to the fore in these works: Instagram has less data than some alternatives (e.g. Twitter), there is an imbalance in classes, and it is hard to get ground-truth labels in the case of complex bots. Consequently, it leads to frequent aggregate data creation by various authors based on several sources or hybrid labeling approaches to enhance the quality of training data. In addition to particular models, systematic research on fraud-detection systems on a large scale indicates the necessity of standard test datasets and standard evaluation. The surveys and systematic reviews suggest the combination of feature types profile, content, temporal and network, as well as the multimedia cross-validation and the adversarial testing, as attackers develop strategies (Liu et al., 2024). [4]. Experimental studies also evidence that a solid balance between interpretability and detection power of feature engineering, classical ML (such as Random Forest, SVM, XGBoost) and deep networks (such as GRU/LSTM/CNN) is achieved when using hybrid frameworks.

Gap Analysis & Applicability to the Current Research.

The current body of literature indicates that (a) only the features of the metadata that are non-dynamic (i.e., follower/following ratio, bio length, tokens in the username) are highly informative, (b) the behavioral/temporal features make them robust and (c) the hybrid deep models tend to perform the best but they need to be trained extremely attentively and the data quality is imperative. Nonetheless, a smaller number of works provide deployable, user accessible systems that integrate various model families, and deliver transparency (model-wise outputs) via a lightweight web interface. The proposed project addresses that gap by (i) using small but significant feature set like many studies in Instagram-detection, (ii) comparing classical with hybrid/deep models (RF, SVM, ANN, GRU, LSTM, Hybrid), and (iii) providing some real-time prediction and visualization in a Flask web app - both in terms of detectives and workability.

*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 13 Issue XII Dec 2025- Available at www.ijraset.com*

## IV. METHODOLOGY

The research methodology adopted will ensure the creation of an end-to-end system that can identify fake Instagram profiles based on machine learning and deep learning models. This can be broken down into five key processes; data preparation, feature extraction, model training, system implementation, and real-time prediction.

### *A. Data Collection*

The data is composed of Instagram users profiles labeled real or fake. The records consist of numeric and behavioral features including the number of followers, the number of following, the number of posts, the content of the bio, the pattern of the usernames, the availability of the profile picture, and the privacy status of the account. The choice of these features is based on the fact that they are frequently utilized to spot suspicious accounts on social media.

### *B. Feature Engineering*

Every Instagram profile is transformed into a numerical feature feature vectors.
Available features are obtained and calculated:
• Follower–Following Ratio
Signifies account influence and a natural growth. ratio = number of followers / following ( max(1,following) )

• Biography Length
The total count of characters in the description of the account.

• Number of Posts (Media Count)
Reports the level of activity of the profile.

• Profile Picture Visibility.
Binary value (1 = present, 0 = absent).

• Private/Public Status
Test value of zero to one to differentiate type of account.

• Username Length
Username length that was cleaned.

•Username Digit Count
Unnatural usernames are often represented by numerical characters.

• Emoji Removal
To eliminate noise caused by special-character usernames, the de Emojify () function of the app.py is used to cleaning the usernames.

Every feature is bundled to a model value which is applied in training and forecasting the model.

### *C. Data Preprocessing*

The dataset is preprocessed to describe it as appropriate to the ML/DL algorithms prior to training:
*1)* Label Encoding target variable (is_fake)
*2)* Models Scaling/Normalization of ANN and SVM.
*3)* Train-Test Split (Mostly 80 training and 20 testing).
*4)* RNN-based deep learning models (GRU, LSTM, Hybrid) Feature Reshaping.
These measures make the data consistent and compatible with the interrogating model networks.

*D. Model Development*

Machine learning and deep learning models are both applied to obtain strong and solid detection.

*1) Machine Learning Models*

- Random Forest Classifier: Thousands of trees are used to define more broadly.
- Support Vector Machine (SVM): Ruby boundary finding (RBF) X applies the RBF kernel to non-linear separation.
- Artificial Neural Network (ANN): Three hidden layers (100 neurons each), which are trained in 2000 steps.

The results of these models give classification results of whether the account is own or fake.

*2) Deep Learning Models*

Loaded using saved .h5 models:

- GRU Model
- LSTM Model
- Hybrid Model

*E. Forms Saving and integration.*

The trained models are all saved and loaded on request:

*1)* ML models saved in .pkl format
*2)* DL models saved in .h5 format

*F. Web Application Development.*

It is worked out using flask and where:

*1)* The data of the Instagram profiles is keyed in by the uses in an HTML form.
*2)* The extracted features are the based features and are transferred to all the trained models through the backend.
*3)* The predictions are processed with the help of a liveprediction) and liveprediction_dl() functions.
*4)* The help of Matplotlib is used to create the comparison graph..
*5)* The output (Real/Fake) is displayed to the user and the outcome of the model displayed in a hassle free manner.

*G. Prediction real time Workflow*

*1)* User involves filling in information on the web-form.
*2)* Flask also characterizes the values and uses these values to create a feature vector.
*3)* All of the models (ML and DL) could be predicted simultaneously.
*4)* Outputs are collected and compared.
*5)* Bar graph is drawn up to observe trend of differences in predictions.
*6)* When the people are shown a result page, the results of the final decision (Real/Fake) are presented.

*H. Model Evaluation*

The tools used to evaluate the trained models include:

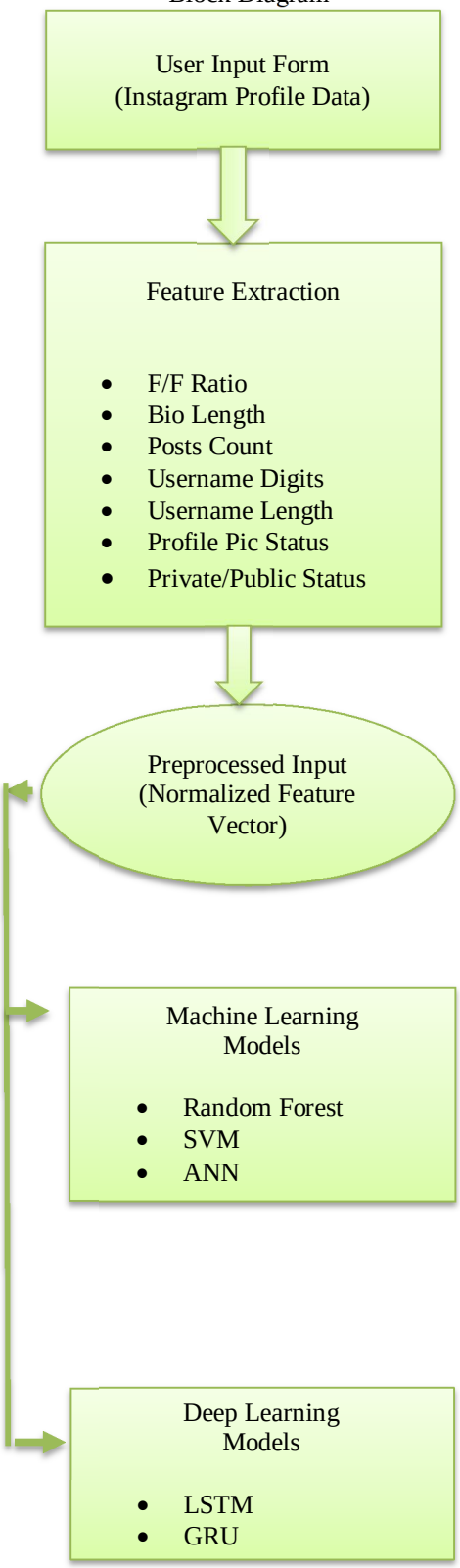
*1)* Accuracy
*2)* Confusion Matrix
*3)* Precision, Recall, F1 Score

Summary

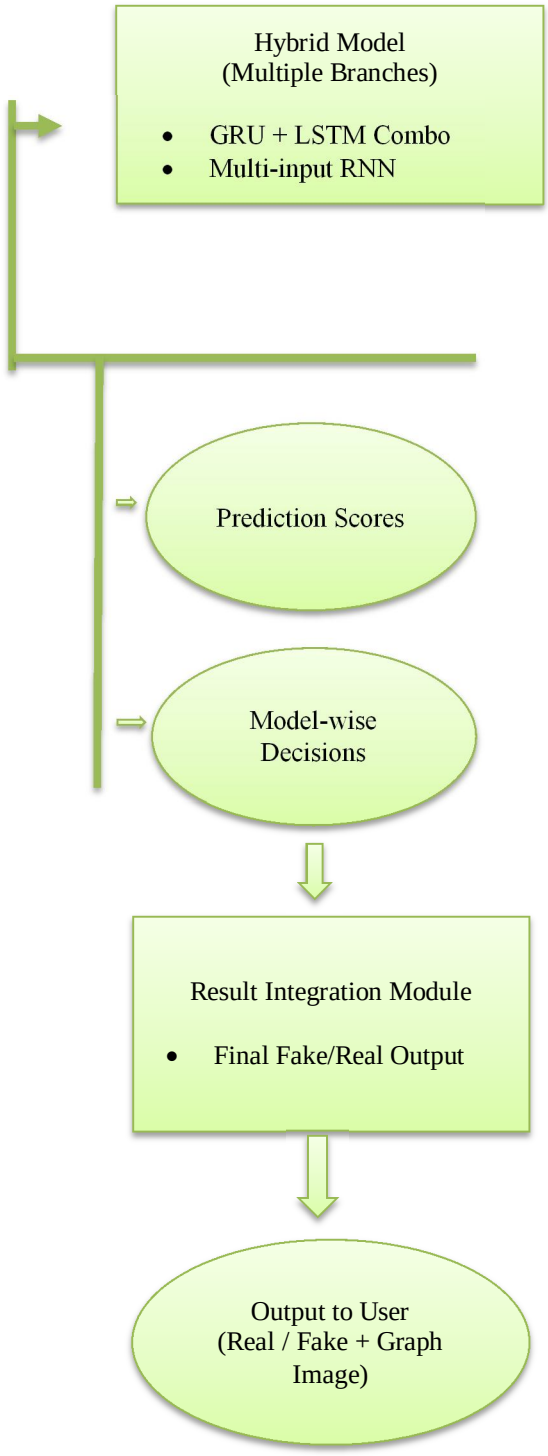
This study adheres to end-to-end paradigm in order to identify fake Instagram accounts on the basis of machine and deep learning solutions. The data is comprised of real and fake Instagram accounts whose features are follower to following ratio, bio length, number of media, and name structure, profile picture, and account privacy. Such characteristics are obtained, processed and turned into numerical vectors to train models. Several models are trained and tested, such as Random Forest, SVM, ANN, GRU, LSTM, and a Hybrid model, to define profiles as trained and tested to define profiles as real or fake.

All the trained models are combined in a Flask-based web application that makes real-time predictions and provides results in relation to models alongside visual comparison graphs. Such an approach will guarantee a realistic, automated, and precise mechanism to identify suspicious Instagram accounts.

Block Diagram



*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 13 Issue XII Dec 2025- Available at www.ijraset.com*




## V. RESULT

Various machine learning and deep learning models have been used to measure the performance of the fake-profile detection system. All were trained on the same set of features to understand how well they could identify real Instagram profiles versus the fake ones by comparing their predictions.

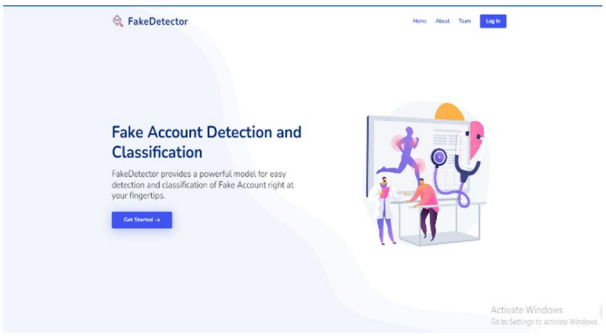



Figure 1 The primary objective of the Instagram Fake Account Detection Site is to help facilitate the efforts of an individual or organisation in determining whether or not a persona is authentic or created for the purposes of misleading others. The overall design approach utilized in the home page layout emphasises clarity, simplicity and intuitiveness for users to easily navigate through and locate the appropriate information or to initiate any requisite tasks. The layout does this by way of presenting primary navigation items (links) along the upper portion of the home page, which provide multiple opportunity point access for the user experience. The home page contains an illustrated image that exemplifies digital analytical processing followed by automated detection.

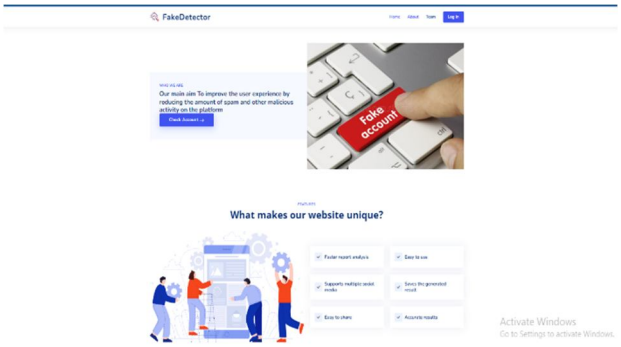



Figure 2. The primary function of this system's About Page is to share information primarily about improving safety for social media users by reducing the volume of spam and other forms of malicious behavior within social media platforms. The layout is neat and clear, providing an easy message for the user and an option "Check Account" for all visitors interested in using the system's detection features. The follower page also highlights some of the advantages of this site, such as quick results from the analysis, simple to navigate and accurate responses.

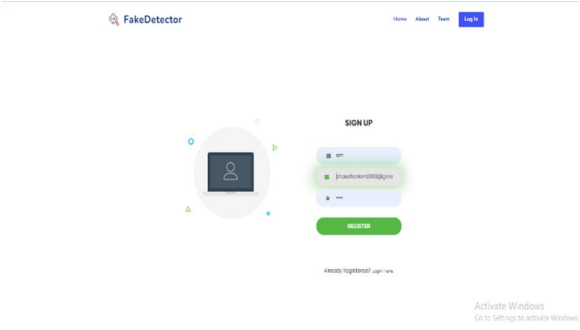



Figure 3. By visiting the Register page, you'll find an uncluttered and straightforward experience when creating your new account. The Register form captures essential name, email, and password fields and provides a single clear "Register" button to complete the process. Additionally, the Register form contains a link pointing to the Login page for those who have previously registered, ensuring that all users can navigate easily between pages and get all the help they need.

*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 13 Issue XII Dec 2025- Available at www.ijraset.com*

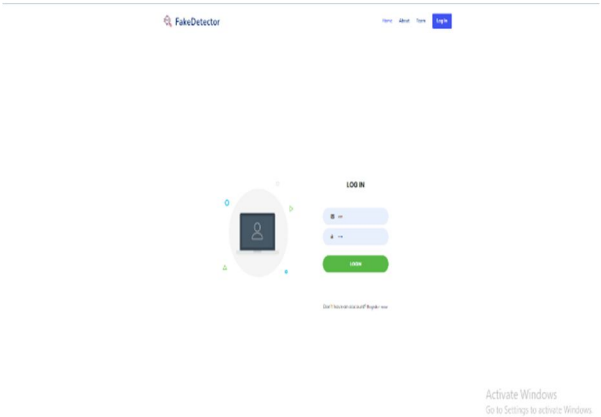



Figure 4 On the Login section, users who are members of this website can log in securely to access their account. They will see fields to enter their email and password, as well as a login button that says "Log In." There will often be a short message below the Login Form to direct first-time visitors to the site on how to register easily.

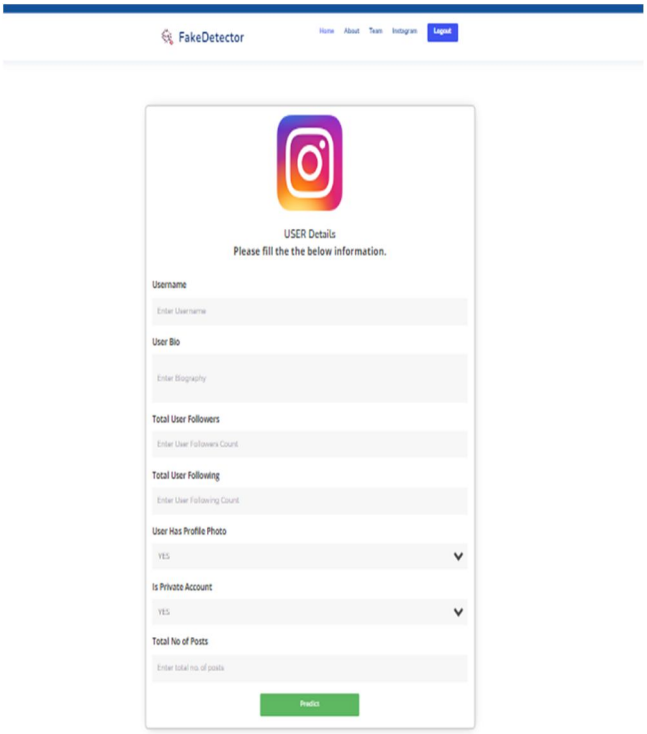



Figure5. Using this Instagram Page, you can easily determine the authenticity of any Instagram account. User information needs to be entered into specific fields which include: the username, bio, follower count, the number of accounts the user follows, whether or not a profile photo is present, the user's profile privacy settings (private or public), and how many posts have been made by the user. Entered into these respective fields is relevant information needed to estimate if the Instagram Account being assessed is real or fake. The result of your findings will be generated after tapping "Predict" at the bottom.

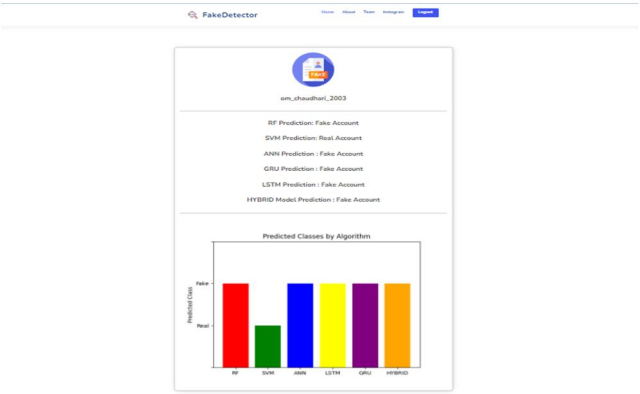



Figure 6. The outcome screen will display the last prediction created for an Instagram account. On the outcome screen, the username is presented at the top followed by the decisions of different machine learning (ML) models of various types (e.g. Random Forest [RF], Support Vector Machines [SVM], Artificial Neural Network [ANN], Long ShortTerm Memory [LSTM], Gated Recurrent Unit [GRU], Hybrid). The majority of the models classify this Instagram account as being fake, while only the SVM ML model has classified the account as real and a bar graph is located at the bottom of the screen and depicts the predictions of the models visually to show how each individual algorithm evaluated the account.

## VI. CONCLUSION

The purpose of this research is to demonstrate how Machine Learning can be applied to determine if a user on Instagram is fake, using different combinations of classifiers. The classifiers include Random Forest, Support Vector Machine (SVM), Artificial Neural Network (ANN), Gated Recurrent Unit (GRU), Long Short Term Memory (LSTM), and Hybrid Classifier. These classifiers work together to provide accurate and reliable results based on various characteristics of Users' Profiles.

Additionally, the research provides a web application that allows Users to check whether their own profile is real or fake. The web application has an easy-to-use interface with the ability to create a profile, log into the web application, view the results from the Classification and check if their profile is real/fake.

In conclusion, this research is an example of how Machine Learning can provide an efficient method for detecting Fake Accounts on Instagram, while simultaneously creating a safer Digital World for all users.

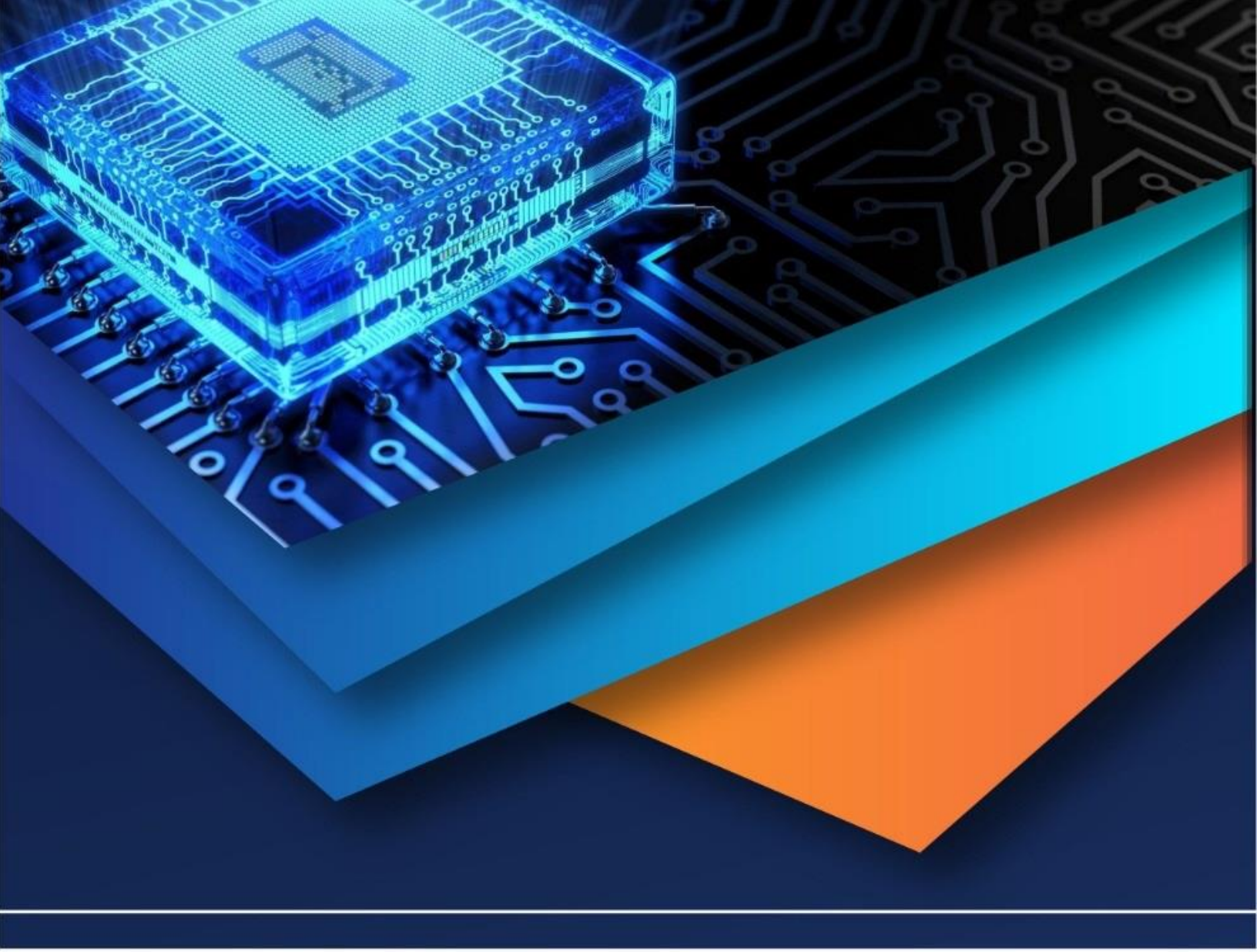
## REFERENCES

[1] Stefanos Chelas, George Routis, Ioanna Roussaki “Detection of Fake Instagram Accounts via Machine Learning Techniques”Submission received: 20 September 2024 / Revised: 6 November 2024 / Accepted: 11 November 2024 / Published: 15 November 2024 https://doi.org/10.3390/computers13110296

[2] Najla Alharbi, Bashayer Alkalifah, Ghaida Alqarawi, Murad A. Rassam “Countering Social Media Cybercrime Using Deep Learning: Instagram Fake Accounts Detection” Submission received: 8 September 2024 / Revised: 28 September 2024 / Accepted: 8 October 2024 / Published: 11 October 2024 https://doi.org/10.3390/fi16100367

[3] Pegah Azami, Pegah Azami, “Detecting Fake Accounts on Instagram Using Machine Learning and Hybrid Optimization Algorithms” Algorithms **2024**, 17(10), 425; **https://doi.org/10.3390/a17100425**

[4] Yunchong Liu, Xiaorui Shen, Yeyubei Zhang, Zhongyan Wang, Yexin Tian, Jianglai Dai & Yuchen Cao “systematic review of machine learning approaches for detecting deceptive activities on social media: methods, challenges, and biases” https://doi.org/10.48550/arXiv.2410.20293



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY