



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJIRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 11, November 2024

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

**Impact Factor: 8.524**

9940 572 462

6381 907 438

[ijirset@gmail.com](mailto:ijirset@gmail.com)

[www.ijirset.com](http://www.ijirset.com)

# Detecting Fake Social Media Profiles

**G.Harsha Vardhan<sup>1</sup>, J.Sathvika<sup>2</sup>, K.Brajesh Rahul<sup>3</sup>, Raja Shekhar<sup>4</sup>**

Department of Computer Science and Engineering, Anurag University, Hyderabad, India<sup>1,2,3,4</sup>

**ABSTRACT:** This project focuses on developing a machine learning-based system to detect fake social media profiles using three algorithms: K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Random Forest. Fake profiles pose a significant threat to online platforms by spreading misinformation, engaging in fraud, and manipulating public opinion. The study aimed to classify profiles as real or fake based on various features such as account age, activity level, and friends count. After training and testing the models, Random Forest achieved the highest accuracy at 93%, followed by SVM at 90% and KNN at 86%. The results indicate that Random Forest is the most effective algorithm for this classification task due to its ability to handle complex data and avoid overfitting. This system has potential applications in improving online security and could be further enhanced with larger datasets and real-time detection capabilities in future research.

## I. INTRODUCTION

Social media platforms have become integral to modern communication, but the growing prevalence of fake profiles poses significant risks to users and platform credibility. These fake accounts are often used for harmful purposes, including scamming, identity theft, and political manipulation. Traditional methods of identifying fake profiles, such as manual reviews or rule-based approaches, are time-consuming and inefficient given the massive scale of social media usage. Therefore, this project focuses on developing a machine learning-based system to automate the detection of fake profiles by analyzing various profile features, such as activity levels, profile completeness, and connections. Using Python, three algorithms—K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Random Forest—were applied to a dataset of social media profiles to determine the most effective model. The results demonstrate the potential of machine learning to enhance detection capabilities, with Random Forest achieving the highest accuracy at 93%, making it a reliable and scalable solution for detecting fake profiles on social media platforms.

## II. RESEARCH METHODOLOGY

### 1 Data Collection

**Data Sources:** Gather datasets containing both genuine and fake social media profiles. In this project, data was collected from:

- **Genuine Users Dataset:** A collection of profiles verified as genuine from various social media platforms.
- **Fake Users Dataset:** A compilation of profiles identified as fake, which may include bots, spam accounts, or impersonators.

**Data Storage:** Store the datasets in CSV format for easy access and manipulation using Pandas in Python

### 2 Data Preprocessing

**Data Loading:** Load the datasets using Pandas and combine them into a single Data Frame for analysis.

**Labelling:** Assign labels to the profiles, with '0' representing genuine accounts and '1' representing fake accounts.

**Handling Missing Values:** Check for and handle any missing values in the datasets using techniques such as:

- Imputation (filling missing values with the mean or median)
- Removal of rows with excessive missing data

**Normalization:** Normalize numerical features (e.g., followers count, statuses count) to ensure that all features contribute equally to the distance calculations in KNN and to improve model performance.

### 3 Feature Extraction and Engineering

**Identifying Relevant Features:** Extract features that are indicative of account authenticity. These may include:

- **User Statistics:** Followers count, statuses count, friends count, favourites count, listed count.
- **Linguistic Features:** Analyse the user's bio and posts for sentiment, length, and word choice, potentially using NLP techniques.

- Profile Characteristics: Account age, the ratio of followers to following, and account type (personal, business).
- Gender Classification: Utilize the NLTK library for gender classification based on the user's name, which can serve as an additional feature.
- Encoding Categorical Variables: Use techniques like Label Encoding to convert categorical variables (e.g., language) into numerical format

#### **4 Model Training**

Splitting the Dataset: Divide the dataset into training and testing sets, typically using an 80/20 split.

Model Selection: Choose KNN and Random Forest classifiers for training:

- K-Nearest Neighbours (KNN): A distance-based algorithm that classifies profiles based on the majority label among the 'k' nearest neighbours.
- Random Forest: An ensemble learning method that builds multiple decision trees and combines their outputs for improved accuracy.
- SVM involves feeding the algorithm labelled training data, where it identifies the optimal hyperplane that separates the classes (fake and real profiles) with the largest margin. During this process, SVM adjusts its parameters to minimize classification errors while maximizing the margin between different classes for accurate predictions.

Training the Models: Fit the KNN, SVM and Random Forest models on the training dataset using the selected features.

#### **5 Model Evaluation**

Performance Metrics: Evaluate the performance of the models using metrics such as:

- Accuracy: The ratio of correctly predicted profiles to the total profiles.
- Precision: The ratio of true positive predictions to the sum of true positives and false positives.
- Recall: The ratio of true positive predictions to the sum of true positives and false negatives.
- F1 Score: The harmonic mean of precision and recall.

Cross-Validation: Implement k-fold cross-validation to ensure the models generalize well to unseen data and to mitigate overfitting.

Confusion Matrix: Generate a confusion matrix to visualize the performance of the classifiers and identify misclassifications.

#### **6 Prediction and User Interface**

User Input: Develop a user-friendly web interface using Flask, allowing users to input profile details (e.g., name, statistics).

Feature Preparation: Preprocess and prepare the input features in the same way as the training data.

Model Prediction: Utilize the trained models to predict whether the input profile is genuine or fake and display the results on the web interface.

#### **7 Results Interpretation**

Analyse and interpret the results of the predictions, discussing the implications of the findings and potential areas for improvement.

Highlight the importance of certain features in the detection process, providing insights into user behaviour and profile characteristics that may indicate authenticity.

#### **8 Future Work**

The future scope of detecting fake social media profiles can expand significantly with the increasing sophistication of fraudulent accounts. In future work, more advanced machine learning techniques, such as deep learning and natural language processing, could be integrated to improve accuracy. Additionally, collecting and analysing larger and more diverse datasets from multiple social media platforms could enhance the system's ability to detect a wider variety of fake profiles. Real-time detection systems could also be developed to identify fake accounts as they are created. Collaboration with social media platforms to implement such systems at scale would provide a more robust solution for online safety and fraud prevention.

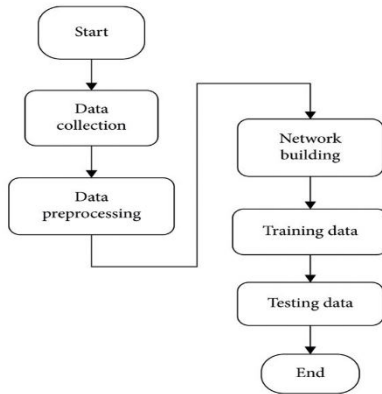


Fig 1: Design of the Project

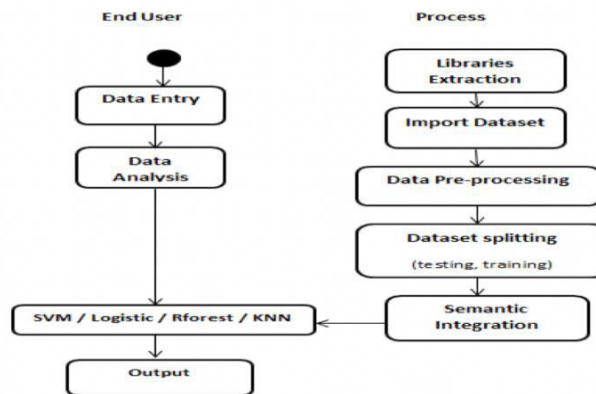


Fig 2: Flow Chart

### III. THEORY AND CALCULATION

The aim of the project is to develop a machine learning-based system that can accurately classify social media profiles as real or fake based on key features extracted from the profile data. Fake profiles on social media platforms are often created for malicious purposes, such as spreading misinformation, phishing, or engaging in fraudulent activities. Detecting these profiles is crucial to maintaining the integrity of social media platforms and protecting users.

In this project, three machine learning algorithms were used: K-Nearest Neighbors (KNN), Support Vector Machine (SVM) and Random Forest (RF). Each algorithm has its own strengths and weaknesses in handling classification tasks. The performance of these algorithms was measured using accuracy as the main metric, and the Random Forest algorithm achieved the highest accuracy.

#### Algorithms Used:

##### 1. K-Nearest Neighbors (KNN):

KNN is a simple, non-parametric algorithm that classifies a data point based on how its neighbors are classified. The algorithm looks at the 'k' nearest points (neighbors) and assigns the most common class among them to the new data point.

It is effective for smaller datasets but can struggle with large datasets as it requires computing the distance to all points in the dataset for classification.

##### 2. Support Vector Machine (SVM):

SVM works by finding the optimal hyperplane that separates the data points into two classes. It is particularly useful for high-dimensional spaces and can work well with both linear and non-linear data using kernel functions.

SVM is known for its robustness in classification tasks but may require fine-tuning for large, noisy datasets.

##### 3. Random Forest (RF):

Random Forest is an ensemble learning method that builds multiple decision trees and merges them to get a more accurate and stable prediction. It reduces overfitting, a common issue with decision trees, by averaging the results of multiple trees trained on different parts of the dataset.

It is highly accurate for complex datasets and was the best-performing model in this study.

#### Calculation for Model Evaluation

##### 1. Data Preprocessing:

The dataset used for this project includes features extracted from social media profiles, such as the number of friends, the frequency of posts, account age, and profile completeness. The data was preprocessed by handling missing values, normalizing numerical features, and encoding categorical variables.

## 2. Model Training and Testing:

The dataset was split into training and testing sets in an 80:20 ratio. The training set was used to build and optimize the models, and the testing set was used to evaluate their performance.

## 3. Accuracy Calculation:

Accuracy was chosen as the performance metric for comparing the three models. Accuracy is calculated as:

$$\text{Accuracy} = \frac{\text{Total Number of Predictions}}{\text{Number of Correct Predictions}} \times 100$$

For each algorithm, the accuracy was calculated using the test dataset:

**KNN Accuracy:** 86%

**SVM Accuracy:** 90%

**Random Forest Accuracy:** 93%

## 4. Confusion Matrix:

To further evaluate the models, confusion matrices were created, showing the number of:

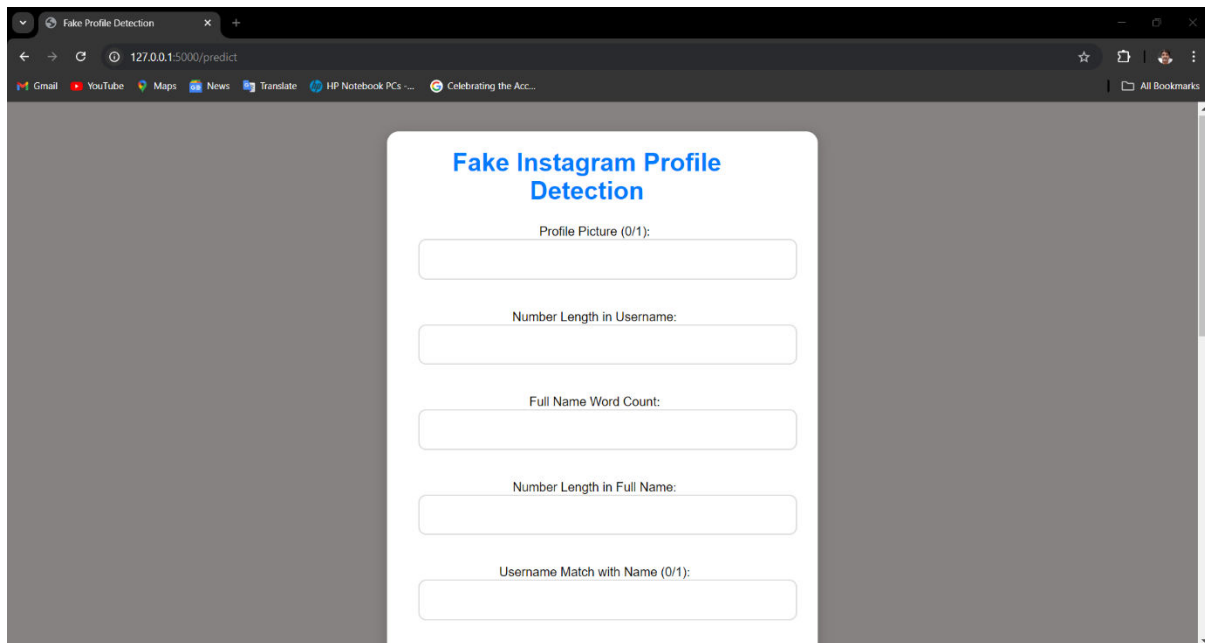
True Positives (TP): Correctly classified real profiles.

True Negatives (TN): Correctly classified fake profiles.

False Positives (FP): Misclassified fake profiles as real.

False Negatives (FN): Misclassified real profiles as fake.

## IV. RESULTS



The screenshot shows a web browser window with the URL `127.0.0.1:5000/predict`. The page title is "Fake Profile Detection". The main content area features a white card with the heading "Fake Instagram Profile Detection". Below the heading, there are five input fields, each with a label and a "0/1" indicator:

- Profile Picture (0/1):
- Number Length in Username:
- Full Name Word Count:
- Number Length in Full Name:
- Username Match with Name (0/1):

Fig 1: Homepage

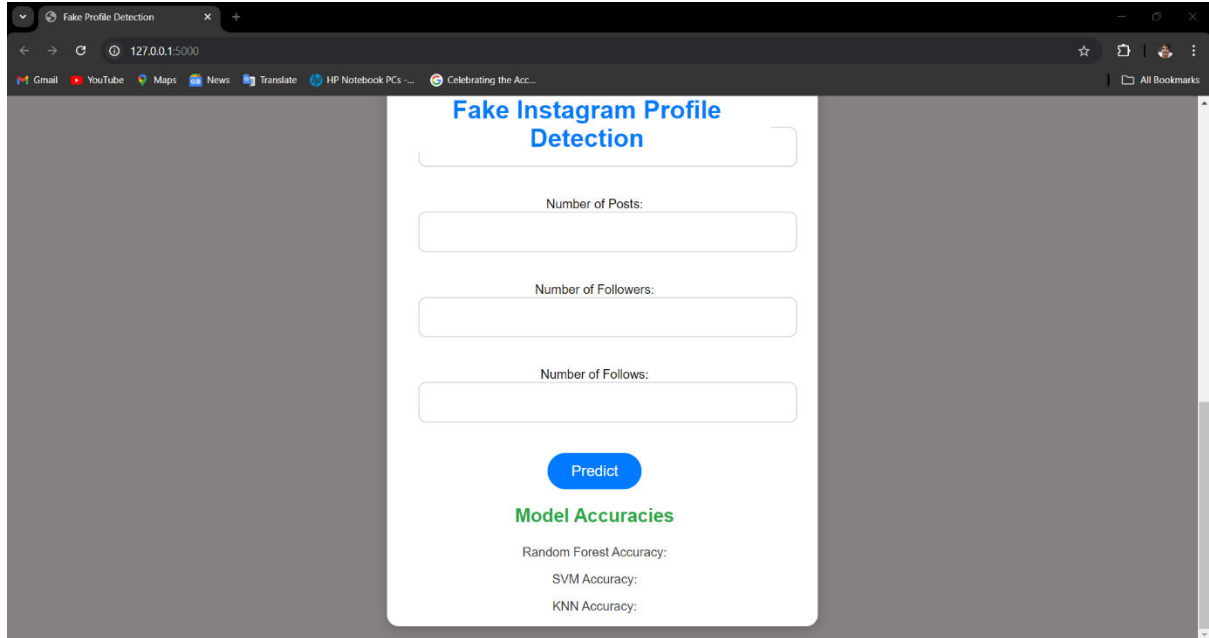


Fig 2: Prediction

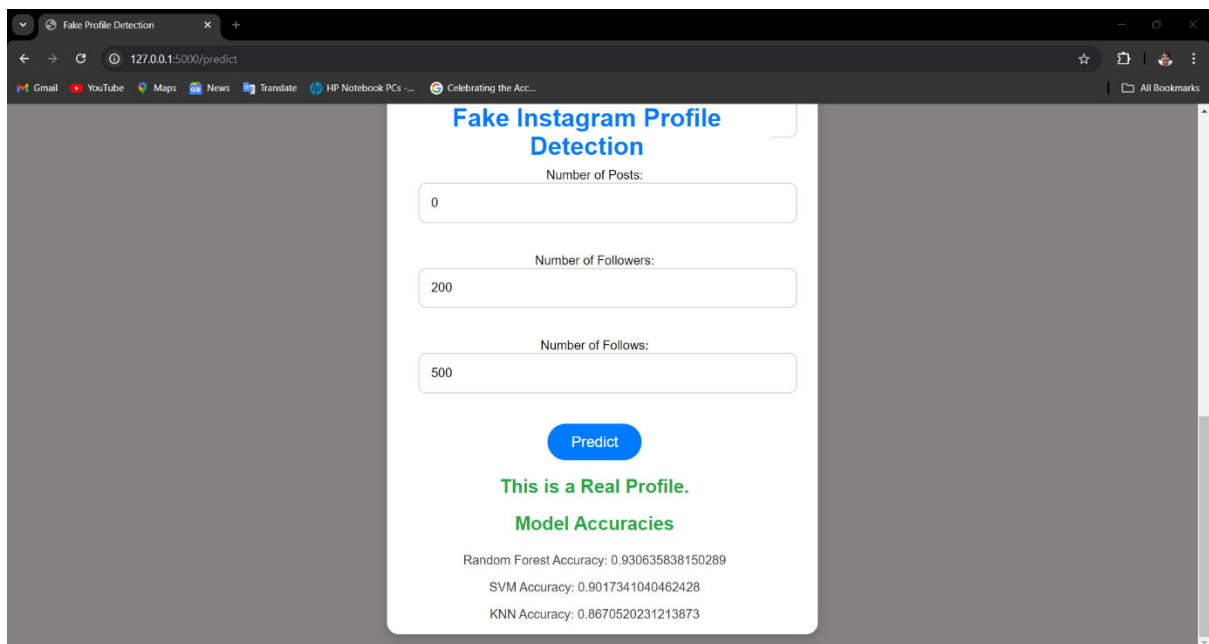


Fig 3: Output for Real Profile

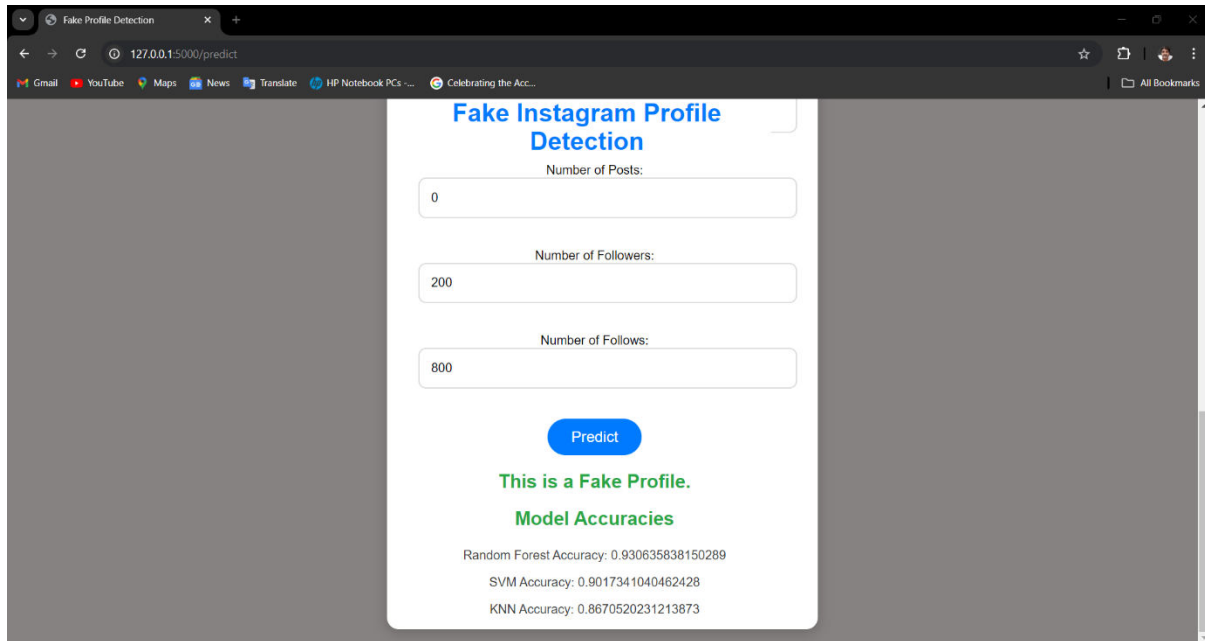


Fig 4: Output for Fake profile

## V. CONCLUSIONS

In conclusion, this project successfully developed a machine learning-based system for detecting fake social media profiles using three different algorithms: K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Random Forest. Among these, the Random Forest algorithm demonstrated the best performance, achieving an accuracy of 93%, making it the most effective model for this task. The results highlight the importance of ensemble learning in handling complex data and detecting patterns that differentiate fake profiles from real ones. While SVM and KNN also performed well, achieving accuracies of 90% and 86%, respectively, Random Forest proved to be the most robust and accurate. Future work can further improve this system by incorporating larger datasets, exploring deep learning techniques, and integrating real-time detection capabilities to enhance the system's efficiency and accuracy across various social media platforms.

## VI. DECLARATIONS

### 1. Study Limitations

One limitation of this study is the potential lack of generalizability due to the dataset used for training and testing the models. If the dataset is not diverse enough, covering only a limited number of social media platforms or user behaviours, the model may not perform well when applied to profiles from other platforms or profiles with different characteristics. Additionally, the features used to detect fake profiles may not capture more sophisticated or evolving techniques used by fake account creators, which could reduce the model's effectiveness over time. The study also focuses on static profile features and does not account for real-time behaviours, which could offer more comprehensive insights into distinguishing real and fake profiles. Finally, there may be a computational limitation, as more advanced machine learning models, such as deep learning, were not explored, which could potentially yield even better results but require significant resources.

### 2. Acknowledgements

I would like to express their sincere gratitude to Anurag University for their unwavering support throughout this project. Special thanks go to Mr G.Rajasekhar, Assistant Professor in the Department of Computer Science and Engineering, for her invaluable guidance and supervision. We also acknowledge the assistance provided by our colleagues and friends for their feedback and encouragement during the development of this project. Their contributions have been instrumental in the successful completion of the "Detecting Fake Social Media Profiles".

#### REFERENCES

- [1]. Aniket Agravat, Umang Makwana, Sahil Mehta, Devashish Mondal, Sushant Gawade “Fake Social Media Profile Detection and Reporting Using Machine Learning” International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), ISSN (Online) 2581-9429 Volume 4, Issue 5, March 2024.
- [2]. K. Harish, R. Naveen Kumar, Dr. J. Briso Becky Bell, " Fake Profile Detection Using Machine Learning”, International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Print ISSN: 2395-1990, Online ISSN: 2394-4099, Volume 10, Issue 2, pp.719-725, March-April-2023. DOI: <https://doi.org/10.32628/ijrsrset2310264>
- [3]. Adarsh Chaurasiya, Amrit Verma, Kota Shamitha, Dr Gaurav Kumar. “Fake Profile Detection in Instagram” International Journal For Research in Applied Science and Engineering Technology. Online ISSN: 2321-9653, Volume 12, Issue 5th May 2024. DOI: <https://doi.org/10.22214/ijraset.2024.60629>
- [4]. Krutika palav, Pranali awari, Siddhi Jiman. “Instagram Fake Account Detection” International Research Journal of Modernization in Engineering and Technology and Science Online ISSN: 2582-5208, Volume 3/Issue 5th may 2021





INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details