International Journal of Scientific Research & Engineering Trends
Volume 10, Issue 5, Sept-Oct-2024, ISSN (Online): 2395-566X

Fake Profile Identification and Classification Using Machine Learning

Professor Disha Nagpure (HOD), Professor Shilpa Shide (Guide) Vaishnavi Gaikwad, Vaishnavi Panchal, Vikrant Kothimbire, Vinay Makwana
Dept. of Artificial Intelligence and Machine Learning
Alard College of Engineering and Management Pune

Abstract- This paper details the design and implementation of Social media platforms are essential for communication today, allowing people to connect, share, and interact. However, the rise of fake profiles on sites like Instagram creates significant challenges related to user privacy, security, and trust. This research proposes a new approach to identify and classify these fake profiles using machine learning techniques. The findings contribute to ongoing efforts to combat fake accounts, promoting a safer and more trustworthy online environment. By leveraging machine learning and a thorough set of features, the model shows promising results in detecting and categorizing fake profiles. This research also opens up opportunities for further exploration, such as integrating different data sources and adapting the model for use on other social media platforms.

Index Terms- Profile Identification, User authentication, Data preprocessing, Model training, Online security, Machine learning

I. INTRODUCTION

Fake profiles on social media can take many forms, from simple bots that post spam to more advanced impersonators who try to trick real users for money, social influence, or other illegal purposes.

Conventional methods, like manual reviews and user reporting, fall short when it comes to handling the massive volume of profiles and interactions online.

This gap highlights the need for more sophisticated technological solutions. Machine learning has emerged as an effective approach for dealing with fake accounts on social platforms. By utilizing machine learning algorithms, we can automatically spot and categorize fake profiles based on unique patterns and characteristics.

The increasing role of social media, the challenges associated with fake accounts, and the progress in machine learning techniques have all contributed to the development of tools designed to identify and classify these profiles. This research seeks to foster a safer and more reliable online space by offering a comprehensive strategy for addressing the issue of fake profiles on Instagram through machine learning.

The research on “Fake Profile Identification and Classification using Machine Learning” is relevant due to its potential to address critical issues related to user trust, online safety, and platform integrity. By leveraging the power of machine learning, this research offers practical solutions that align with the needs of the digital age.

Social networking platforms encounter a variety of issues, such as trolling, harassment, and cyberbullying, alongside the prevalence of fake profiles. These deceptive accounts often represent individuals who use false identities and credentials. While many seek to earn a legitimate income online, the internet can also be misused for profit through fraudulent activities

People create fake profiles for various reasons, including marketing, campaigning, impersonation, and social engineering. To protect users from spam, phishing, and other types of fraud, Facebook has established its own security measures.

Objective

This research seeks to promote a safer online environment, improve user experiences, and aid social media platforms in their fight against fake profiles. It involves identifying and extracting key features from collected Instagram profiles and their associated content. The study also focuses on creating an efficient machine learning model to identify and classify these fake Instagram accounts.

II. LITERATURE SURVEY

Prominent websites on the internet face constant threats from spammers, scammers, and phishers who aim to inundate users

with unwanted spam while stealing their personal information. These attackers have access to vast resources, including global botnets, substantial funding, dedicated personnel, and control over compromised accounts. Protecting our users from such threats presents a significant challenge in adversarial learning, requiring scalable and robust solutions.

To address this issue, we have developed and implemented a real-time system over the past few years that is coherent, scalable, and adaptable. This system monitors every read and write operation, conducting real-time classifications and inspections, functioning like an immune system for the platform. Online impersonation and false identities are prevalent in our daily interactions on social networks.

NO.	Title	Year	Method
1.	Prediction of fake Instagram profiles using machine learning	2023	Chi-Square algorithm, Logistic Regression,Random forest algorithms are used
2.	Detection of fake Account in Instagram using machine learning	2019	Logistic Regression and Random forest Algorithms are used
3.	Fake account detection on social media	2023	SVM, Random Forest, decision Tree, K-Nearest neighbour, Gaussian Naïve Bayes Algorithm
4.	Fake profile detection using machine learning techniques	2022	Extreme Gradient decent, long short term memory algorithms used

In this project, we propose a methodology to determine whether an account is original or fake. Our model employs Support Vector Machine (SVM) as a classification technique, enabling it to process large datasets of accounts simultaneously, which eliminates the need to evaluate each account one by one.

In today's digital landscape, social media dominates various aspects of life, with the number of users growing every day. The rapid increase in social media usage offers enhanced communication with others. However, this also creates new avenues for attacks, such as fake identities and misleading information. Recent studies indicate that the number of social media accounts significantly exceeds the actual number of unique users.

To tackle this issue, we can leverage machine learning algorithms to identify both fraudulent and legitimate accounts. By utilizing various models, these algorithms can analyze and categorize datasets effectively. Sometimes, distinguishing the outcomes of different models can be challenging, so we can streamline this process by adopting a hybrid approach to machine learning techniques.

The research encompasses developing an effective machine learning-based model for identifying and classifying fake social media profiles.

III. METHODOLOGY

1. Data Flow Diagram

A Data Flow Diagram (DFD) illustrates the flow of data within a system. In DFD0, the context diagram, we represent the overall system. A rectangle indicates the external entities (both input and output), while a circle represents the system itself. In DFD1, we break down the inputs and outputs in more detail. For example, in our system, the input could be either text or images, and the output would be whether a rumor is detected. DFD2 further details the operations within the system, outlining the actions of both the user and the admin.

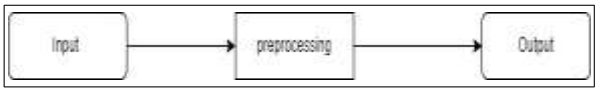

Figure 1: DFD(0) diagram

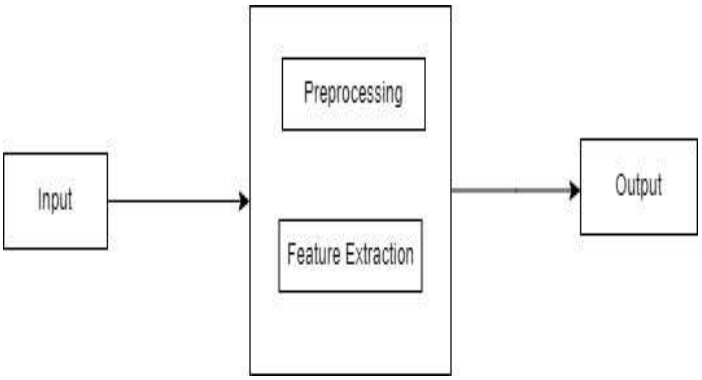

Figure 2: DFD(1) diagram

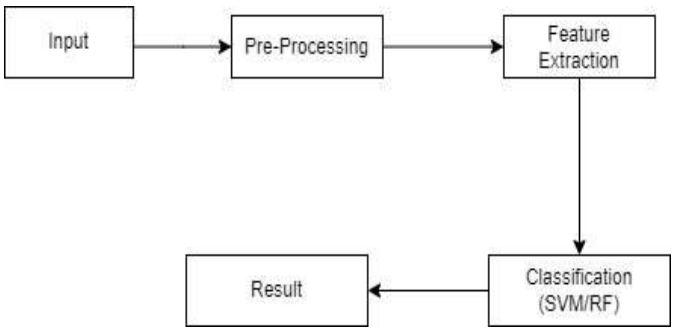

Figure 3: DGD(2)diagram

International Journal of Scientific Research & Engineering Trends
Volume 10, Issue 5, Sept-Oct-2024, ISSN (Online): 2395-566X

2. Flow Chart

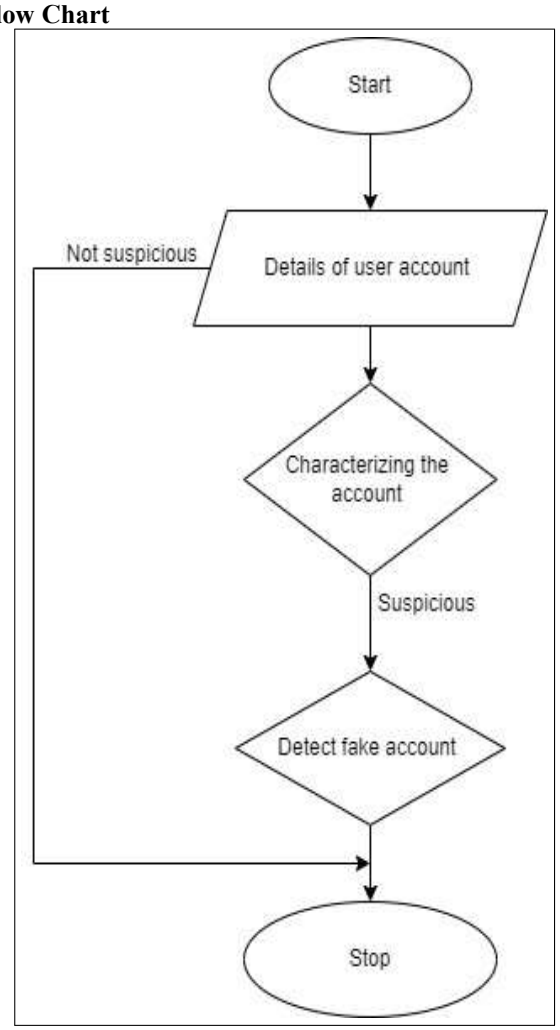

Fig 4: Flow chart

IV. SYSTEM ARCHITECTURE

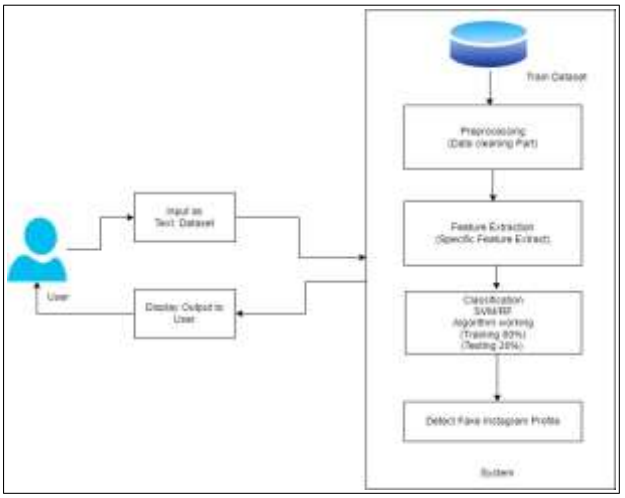

Fig. 5: System architecture

1. Module

Admin

The admin must log in with a valid username and password. Once logged in, the admin can perform various tasks, such as viewing and authorizing users.

View and Authorize Users

The admin can see a list of all registered users and their details, including username, email, and address. The admin has the ability to authorize users.

View Chart Results

The admin can view different data charts, including product search ratios, keyword search results, and product review rankings.

End User Module

This module allows multiple users to register and perform operations. Users need to register before accessing features. Once registered, their information is stored in the database. After successful registration, users must log in with their authorized username and password. Upon successful login, users can manage their accounts and perform other tasks

2. UML Diagrams

Unified Modeling Language (UML) is a standardized way to create software blueprints. It's used to visualize, define, build, and document the components of complex software systems. UML itself is independent of any specific process, but it works best in a process that focuses on use cases, is architecture-cantered, and follows an iterative, step-by-step approach. There are various types of UML diagrams available to represent different aspects of a system.

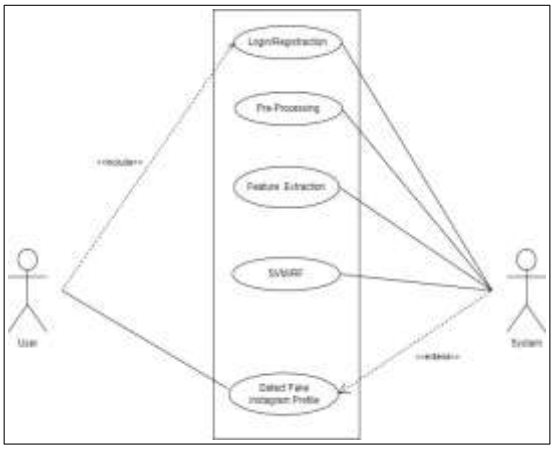

Fig 6: Use case Diagram

3. Sequence Diagram

A sequence diagram is a type of UML diagram that shows how objects in a system interact with each other over time. It

visually represents the sequence of messages exchanged between objects to carry out a specific process or operation.

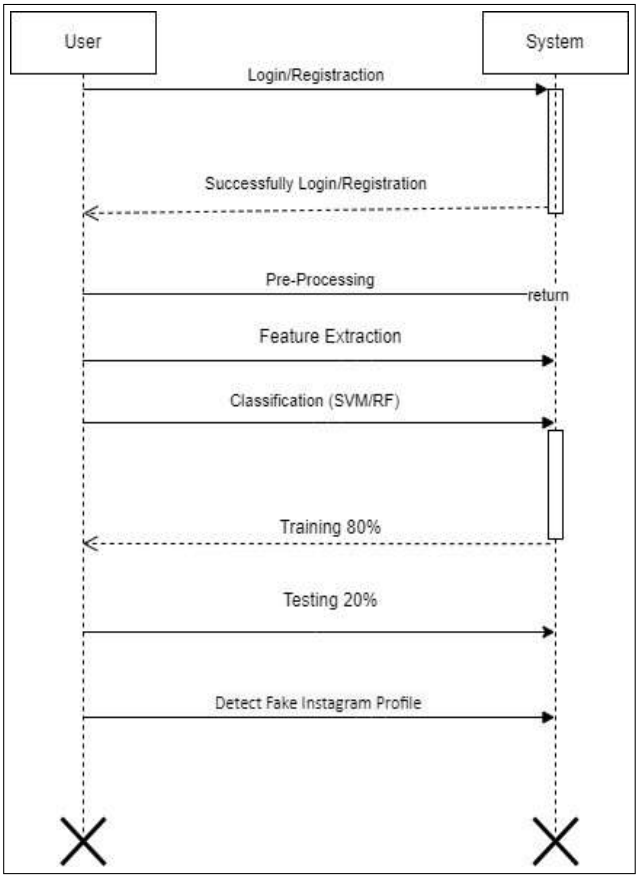

Fig 7: Sequence Diagram

V. ALGORITHMS

1. Support Vector Machines

Support Vector Machines (SVMs) are a popular type of supervised machine learning algorithm used for classification and regression tasks. SVMs can perform both linear and non-linear classification. Support Vector Machine Working Steps:

- Step 1: Load the important libraries.
- Step 2: Import dataset and extract the X variables and Y separately.
- Step 3: Divide the dataset into train and test.
- Step 4: Initializing the SVM classifier model.
- Step 5: Fitting the SVM classifier model.
- Step 6: Coming up with predictions.
- Step 7:Evaluating model's performance.

2. Random Forest

Random Forest is an ensemble learning method that is widely used for both classification and regression tasks in machine learning. Random Forest is an ensemble learning technique that combines the predictions of multiple decision trees to improve the overall predictive accuracy and robustness. Random Forest Working

Steps

- Step 1: Importing and processing the data.
- Step 2: Training the random forest classifier.
- Step 3: Testing the prediction accuracy.
- Step 4: Visualizing the results of the classifier.

3. Mathematical Model
4. Let S be the Whole system S= I,P,O
5. I-input -procedure -output
6. Input(I)
7. I=Text Dataset

Where,

- Dataset- Text dataset Classification Using SVM/RF Algorithm
- Procedure (P),
- P=I, Using I System Detect Fake Instagram Profile.

VI. CONCLUSION

The research on "Fake Profile Identification and Classification using Machine Learning" addresses the ongoing problem of fake accounts on social media, specifically Instagram. This study uses machine learning techniques to enhance user safety and foster a more reliable online environment. By boosting user confidence and preserving the integrity of social media interactions, this research has real-world implications beyond academic circles, benefiting individuals, businesses, and society at large. As social media continues to evolve, this work plays a key role in promoting trust and authenticity, encouraging positive online interactions and collaborations.

REFERENCES

1. Aleksei Romanov, Alexander Semenov, Oleksiy Mazhelis and Jari Veijalainen.2017. "Detection of Fake Profiles in Social Media". In 13th International Conference on Web Information Systems and Technologies.
2. Indira Sen, Anupama Aggarwal,Shiven Mian.2018."Worth its Weight in Likes: Towards Detecting Fake Likes on Instagram". In ACM International Conference on Information and Knowledge Management.Nazir, Atif, Saqib Raza, Chen-Nee Chuah, Burkhard Schipper, and C. A. Davis. "Ghostbusting Facebook: Detecting and Characterizing Phantom Profiles in Online Social Gaming Applications." In WOSN. 2010.
3. Nambouri Sravya, Chavana Sai praneetha, S. Saraswathi," Identify the Human or Bots Twitter Data using Machine Learning Algorithms", International Research Journal of Engineering and Technology

(IRJET), Volume: 06 Issue: 03 — Mar 2019 www.irjet.net, e-ISSN: 2395-0056, p- ISSN: 2395-0072.

4. M. Smruthi, N. Harini," A Hybrid Scheme for Detecting Fake Accounts in Facebook", International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7, Issue-5S3, February 2019.
5. Nazir, Atif, Saqib Raza, Chen-Nee Chuah, Burkhard Schipper, and C. A. Davis. "Ghostbusting Facebook: Detecting and Characterizing Phantom Profiles in Online Social Gaming Applications." In WOSN. 2010.
6. Rao, P. S., J. Gyani, and G. Narsimha. "Fake profiles identification in online social networks using machine learning and NLP." Int. J. Appl. Eng. Res 13.6 (2018): 973-4562.