Safeguarding social media: Fake profile detection on Instagram using ML

Prof. Disha Nagpure (HOD), Prof. Dr. 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: **Social media has become a major way for people to communicate, share content, and stay connected. However, the rise of fake accounts on platforms like Instagram is a growing concern. These fake profiles can spread misinformation, scam users, and harm online communities. This project aims to detect fake profiles using machine learning techniques. We use three popular algorithms—Support Vector Machine (SVM), Random Forest, and Decision Tree—to classify user accounts as either real or fake. The models are trained on a dataset that includes user activity, engagement levels, and content features.**

Our main goal is to compare how well each model can identify fake accounts. The results show that machine learning can be a powerful tool in improving safety and trust on social media. This research also suggests future directions, such as combining more data types and applying the system to other platforms.

Keywords: **Fake accounts, Social media, Machine learning, SVM, Random Forest, Decision Tree, Online safety, User behavior, Data analysis.**

1. INTRODUCTION

In today`s Modern society, social media performs a essential function in everyone's life. The preferred motive of social media is to hold in contact with friends, sharing news, etc. The range of customers in social media is growing exponentially. Instagram has lately received significant reputation amongst social media customers. With greater than 1 Billion energetic customers, Instagram has turn out to be one of the maximum used social media sites. After the emergence of Instagram to the social media scenario, human beings with a very good range of fans were referred to as Social Media Influencers. These social media influencers have now turn out to be a go-to area for the commercial enterprise employer to market it their merchandise and services[1].The good sized use of social media has turn out to be each a boon and a bane for the society. Using Social media for on line fraud, spreading False records is growing at a speedy pace.[2][3]

Social media plays a vital role in modern life, with Instagram standing out as one of the most popular platforms, hosting over a billion users. The platform's growth has led to the rise of influencers, who are often central to digital marketing efforts.

However, this popularity has also brought challenges—particularly the spread of fake profiles used for scams, misinformation, and other harmful activities. These accounts threaten user safety and platform integrity.

This research proposes a machine learning-based system to detect fake Instagram profiles. It includes a user registration system backed by an SQLite database, which helps analyze and compare user data against known patterns of fake behavior.

By applying Support Vector Machine (SVM), Random Forest, and Decision Tree algorithms, the study compares model performance in identifying fake accounts, aiming to support safer and more trustworthy social media spaces.

II. RELATED WORK

With the growing number of fake accounts on social media, detecting such profiles has become a major focus in recent research. Early studies used basic machine learning models like Logistic Regression, Decision Trees, and Support Vector Machines (SVM) to find patterns in user behavior, such as follower-following ratios, posting habits, and engagement levels. For example, Dey et al. demonstrated the effectiveness of Decision Trees and SVM in detecting fake Instagram profiles using profile-based features.

Other studies explored text-based features by analyzing bio descriptions, captions, and comments. Natural Language Processing (NLP) and sentiment analysis were applied to evaluate user-generated content and detect signs of fake activity. Some researchers also looked at social network behavior—like interaction patterns and message flows—viewing the user within a larger network structure to catch unusual or automated behavior. Akyon and Kalfaoglu proposed a method that combined behavior data with metadata to identify bots on Instagram.[12]

While these approaches showed promising results, many were limited to specific feature types or datasets. In contrast, our research takes a broader approach by combining both behavior-based and content-based features. We also compare several machine learning models—SVM, Random Forest, and Logistic Regression—to find the most effective solution for detecting fake accounts in a scalable and accurate way.

III.METHODOLOGY

This study employs a structured machine learning approach to identify and classify fake Instagram profiles. The methodology consists of four key phases: (1) Data collection of profile metadata and activity patterns, 2) Data Preprocessing (3) Feature extraction focusing on behavioral, textual, and network characteristics,

(4) Model development using SVM and Random Forest algorithms, and (5) Performance evaluation through standard classification metrics. The systematic pipeline ensures reliable detection while maintaining adaptability to evolving fake profile tactics.

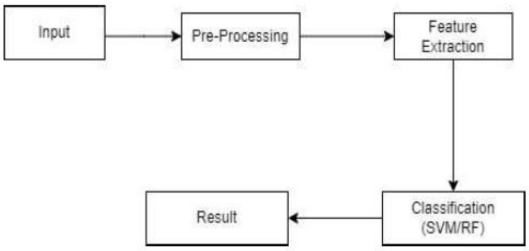


Fig 1 – Data Flow Diagram

Data Collection

To create an effective fake profile detection system, we gathered a large dataset that includes both real and fake Instagram profiles. These profiles were either labeled manually or obtained from reliable, pre-verified datasets. The data collected includes key details like the number of followers, followings, posts, account creation date, bio content, presence of profile pictures, and user interaction frequency.

Data Preprocessing

Once the raw data was collected, it was cleaned and processed to fix inconsistencies and standardize values. For example, numerical features like the ratio of followers to followings and average likes per post were calculated. Textual data, such as the bio description, was analyzed to look at keyword usage, bio length, and the presence of emojis or hashtags. We also examined behavioral patterns like how often users post and how consistently they engage with others to further enhance the feature set.

Feature Extraction

The next step involved extracting meaningful features from the collected data. We looked at both the content (like bio and posts) and behavior (such as interaction frequency and engagement rates) to identify characteristics that can help distinguish real profiles from fake ones.

Model Selection and Training

We tested several supervised learning models, including Random Forest, Support Vector Machine (SVM), and Logistic Regression, to determine which one works best for identifying fake profiles. The data was divided into training (80%) and testing (20%) sets. To ensure the models didn't overfit to the training data, we used a 10-fold cross-validation method. We also fine-tuned the model’s settings (called hyperparameters) using a grid search to get the best performance.

Performance Evaluation

To evaluate how well the models performed, we used various metrics such as accuracy, precision, recall, and F1-score. These metrics help us understand how well the models can correctly identify fake profiles, minimize errors, and maintain reliable performance across different scenarios.

IV ALGORITHMS

A. Support Vector Machine (SVM)

Support Vector Machine (SVM) is a method used to classify data by finding the best line (or hyperplane) that divides the data into two different categories. In this case, it helps separate real profiles from fake ones. SVM tries to create the widest possible gap (margin) between the two groups, making it easier to classify new data correctly.

SVM is particularly useful when the data cannot be separated with a simple straight line. It uses a technique called the "kernel trick" to transform the data into higher dimensions, allowing for a clearer separation between the classes.

In this project, SVM is used to distinguish between real and fake Instagram profiles. By analyzing various features like bio text, follower numbers, and engagement levels, SVM finds a clear dividing line between authentic and fraudulent accounts.

Once trained, the model can predict whether a new Instagram profile is more likely to be real or fake based on its characteristics. This makes SVM a powerful tool for spotting subtle differences that might indicate fake accounts.

B. Random Forest (RF)

Random Forest is a machine learning method that works by creating many decision trees instead of just one. Each tree is trained using different parts of the data. When it needs to make a prediction, it checks what most of the trees suggest and goes with the majority vote.

Each tree looks at things like how many followers a profile has, how often it posts, how people interact with it, and what's written in the bio. By using many trees and combining their results, the model becomes more accurate and avoids mistakes that a single tree might make.

In this project, Random Forest helps decide whether an Instagram account is real or fake. It does this by looking at different details from the profile and comparing them to known patterns of real and fake accounts.

Because it uses the results from many trees, it's better at spotting fake profiles and gives more reliable predictions.

C. Decision Tree (DT)

A Decision Tree is a model that makes decisions by asking a series of questions about the data—kind of like a flowchart. At each step (called a node), it checks a feature (like number of posts or followers) and then splits the data based on the answer. This continues until it reaches a final decision—whether the profile is real or fake.

The tree is built by choosing the best questions (features) at each step to separate the data in the most accurate way. It keeps splitting until it can make a confident decision. One of the best things about Decision Trees is that they're easy to understand and follow.

In this project, the Decision Tree helps figure out if an Instagram profile is genuine or fake. It uses details like how often the user posts, their follower count, and how much engagement they get.

The model learns from examples of real and fake accounts, then uses that knowledge to check new profiles. Because of its clear structure, it's useful for spotting patterns linked to fake behavior.

V. DATABASE STORAGE

To manage user data securely and efficiently, we implemented a lightweight, file-based SQLite database. SQLite was chosen for its simplicity, portability, and integration ease with Python-based data processing scripts. It allowed for local data persistence and effective user credential management during the model development phase.

The database schema included tables such as user credentials, ensuring a normalized and organized structure. Sensitive data such as login information, Registration Form and User's Password was encrypted using standard cryptographic libraries to prevent unauthorized access.

VI SYSTEM ARCHITECTURE

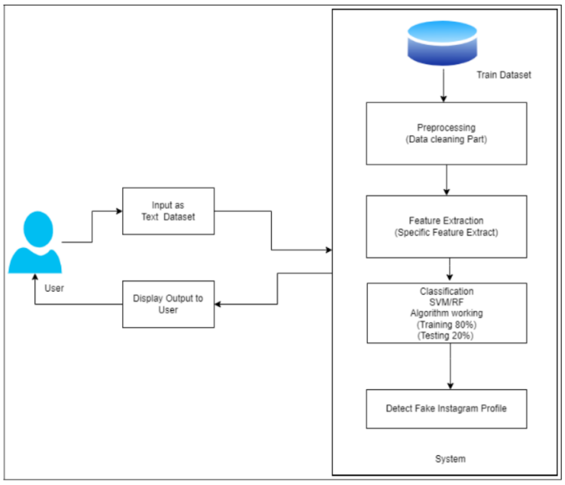
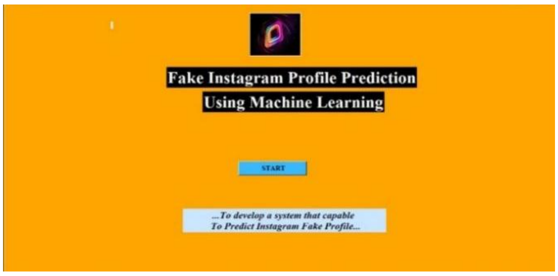


Fig 2– System Architecture

VII. RESULT ANALYSIS


Fake Instagram Profile Prediction
Using Machine Learning

...To develop a system that capable
To Predict Instagram Fake Profile...

Fig 3. System GUI

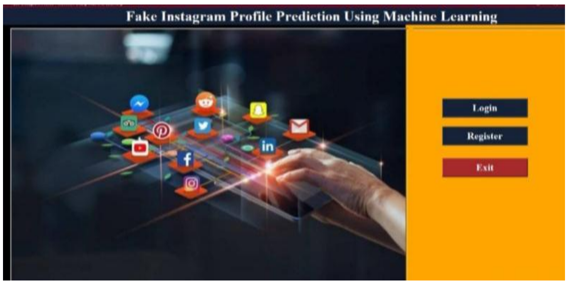
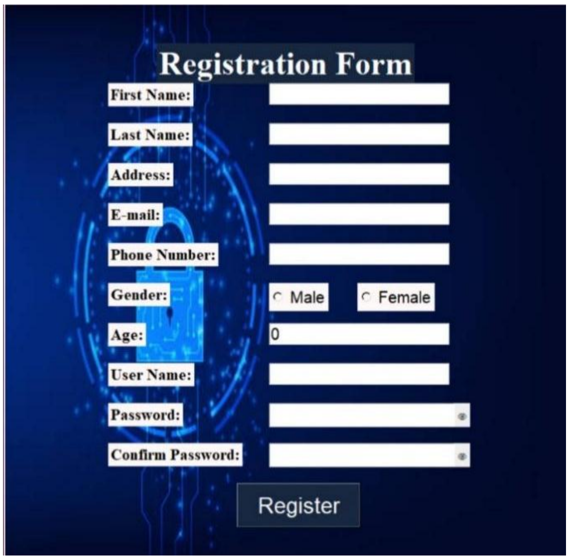


Fig 4. Login Page



Registration Form

First Name:

Last Name:

Address:

E-mail:

Phone Number:

Gender: ○ Male ○ Female

Age: 0

User Name:

Password:

Confirm Password:

Fig 5. Registration Form

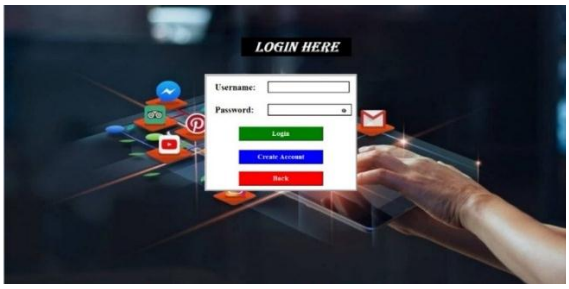


Fig 6. Login

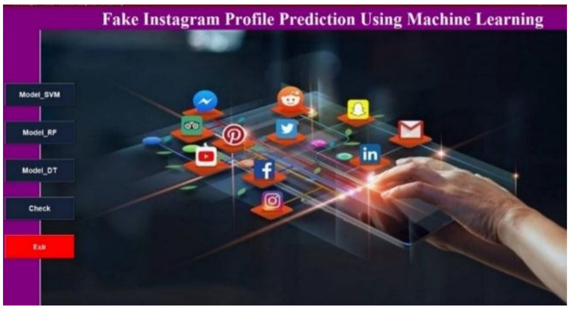
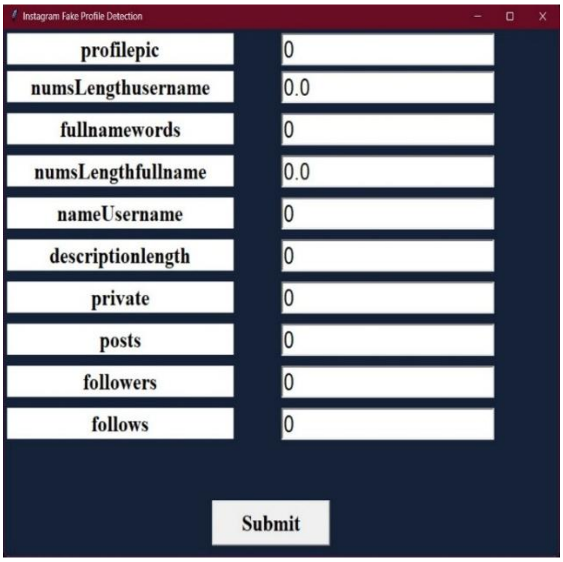


Fig 7. Result Analysis



profilepic	0
numsLengthusername	0.0
fullnamewords	0
numsLengthfullname	0.0
nameUsername	0
descriptionlength	0
private	0
posts	0
followers	0
follows	0

Fig 8. Fake Profile Check

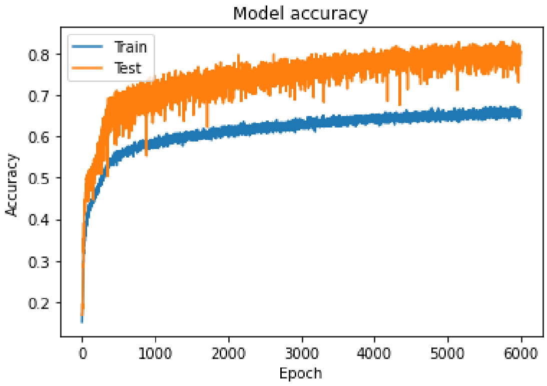


Fig 9. Model Accuracy

VIII. FUTURE IMPROVEMENT

- Deep Learning: Use BERT for advanced text analysis.
- Real-Time Detection: Deploy as an API for live monitoring.

IX. CONCLUSION

This paper presented a machine learning-based approach for the identification and classification of fake Instagram profiles. The proposed methodology contributes to enhancing platform security by accurately distinguishing between genuine and fraudulent accounts. The results demonstrate the practical significance of such systems in promoting user trust and maintaining the integrity of social interactions online. As social media continues to grow, the implementation of intelligent detection techniques will be vital in ensuring a safe and authentic user experience.

This study proposed a machine learning-driven framework for detecting and classifying fake accounts on Instagram, addressing a critical concern in modern social media usage. By applying supervised learning algorithms to behavioral and content-based features, the system effectively distinguishes between legitimate and fraudulent profiles. The experimental results support the feasibility and accuracy of the approach, showcasing its potential for real-world application.
The significance of this work extends beyond academic exploration. With the increasing influence of social media on communication, marketing, and public opinion, the ability to identify deceptive users is essential for ensuring user safety and preserving platform integrity. Businesses, influencers, and regular users alike can benefit from a system that reduces misinformation, prevents manipulation, and fosters trust.

Furthermore, this research lays the groundwork for future enhancements, including the integration of deep learning techniques, real-time detection mechanisms, and cross-platform fake account identification. As digital platforms continue to evolve, the development of robust, intelligent security measures will be crucial in sustaining a healthy and authentic online ecosystem.

X. REFERENCES

[1] A. M. Vegni, V. Loscri, A. Benslimane, SOLVER: A Framework for the Integration of Online Social Networks with Vehicular Social Networks, in: IEEE Network 34(1), 2020,pp.204-213,doi: 10.1109/MNET.001.1900259. https://ieeexplore.ieee.org/document/8839970

[2] Ml-cheatsheet.readthedocs.io. (2019). Logistic Regression — ML Cheatsheet documentation. [Online] Available at: https://mlcheatsheet.readthedocs.io/en/latest/logisticregression.html#binarylogistic-regression [Accessed 10 Jun. 2019]

[3] A. U. Hassan, et al., Sentiment analysis of social networking sites (SNS) data using machine learning approach for the measurement of depression, in: 2017 International Conference on Information and Communication Technology Convergence (ICTC), Jeju,2017,pp.138140,doi:10.1109/ICTC.2017.8190959 https://ieeexplore.ieee.org/document/8190959

[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 Pro- files in Online Social Gaming Applications.” In WOSN. 2010.

[6] M. S. Kumar, J. Sabeena, K. M. Veena, K. Pavan, M. Sukavya and K. Sravanthi, "Fake Profile Detection on Social Networking Websites using Machine Learning," 2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS), Coimbatore, India, 2023, pp. 119-122, doi: 10.1109/ICSCSS57650.2023.10169168. keywords: {Machine learning algorithms; Social networking (online);Time series analysis; Boosting; Natural language processing; Security; Decision trees; Fake profiles; Machine learning methods; Natural Language Processing (NLP);Timestamp; Extreme Gradient Boosting algorithm}, https://ieeexplore.ieee.org/document/10169168

[7] Harish, K. & Kumar, R. & Bell J, Briso Becky. (2023). Fake Profile Detection Using Machine Learning. International Journal of Scientific Research in Science, Engineering and Technology. 719-725. 10.32628/IJSRSET2310264. https://www.researchgate.net/publication/370484393_Fake_Profile_Detection_Using_Machine_Learning

[8] Puli, Sreekanth. (2023). Detection of Fake Social Media Profiles Using Machine Learning Techniques. https://www.researchgate.net/publication/373142154_Detection_of_Fake_Social_Media_Profiles_Using_Machine_Learning_Techniques

[9] M. Santhoshi, S. Sailaja and J. Jyotsna, "Deep Learning Approach for Identification of Fake Profiles in Social Media," 2023 World Conference on Communication & Computing (WCONF), RAIPUR, India, 2023, pp. 1-7, doi: 10.1109/WCONF58270.2023.10235178. keywords: {Support vector machines;Deep learning; Training; Machine learning

algorithms; Computational modeling; Neural networks; Predictive models; SVM; random forest; neural networks; fake accounts; machine learning; social networks}, https://ieeexplore.ieee.org/document/10235178

[10] Ezarfelix, Juandreas & Jeffrey, Nathannael & Sari, Novita. (2022). Systematic Literature Review: Instagram Fake Account Detection Based on Machine Learning. Engineering, MAthematics and Computer Science (EMACS) Journal. 4. 25-31. 10.21512/emacsjournal.v4i1.8076. https://www.researchgate.net/publication/358590043_Systematic_Literature_Review_Instagram_Fake_Account_Detection_Based_on_Machine_Learning

[11] F. C. Akyon and M. Esat Kalfaoglu, "Instagram Fake and Automated Account Detection," 2019 Innovations in Intelligent Systems and Applications Conference (ASYU), Izmir, Turkey, 2019, pp. 1-7, doi: 10.1109/ASYU48272.2019.8946437. keywords: {Media; Twitter; Measurement; Tagging; Feature extraction; fake engagement; machine learning; online social networks; Instagram; genetic algorithm; smote}, https://ieeexplore.ieee.org/document/8946437?utm_source=chatgpt.com

[12] Akyon, F. C., & Kalfaoglu, M. E. (2019). Instagram Fake and Automated Account Detection. *2019 Innovations in Intelligent Systems and Applications Conference (ASYU)*. https://export.arxiv.org/abs/1910.03090?utm_source=chatgpt.com