FAKE USER PROFILE IDENTIFICATION SYSTEM USING ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

Chaitanya Chugh, Chayan Pandey, Ishan Saraf, Lakshay Panwar

Department of Computer Science and Engineering

Inderprastha Engineering College, Sahibabad, Ghaziabad, India

Ms. Harshita

Assistant Professor

Department of Computer Science and Engineering

Inderprastha Engineering College, Sahibabad, Ghaziabad, India

Abstract-- **The main theme of our paper is identifying whether the Instagram profile is genuine or fake. Algorithms will be trained with all previous users fake and genuine account data and then whenever we give new test data then trained model will be applied on new test data to identify whether given new account details are from genuine or fake users. The machine learning-based methods were used to perceive false accounts that could give the wrong impression about people. The dataset is pre-processed using a variety of python libraries and a comparison form is obtained to get a realistic algorithm appropriate for the specified dataset. An effort to notice forged accounts on the social media platforms is strong-minded by a variety of machine learning algorithms. The performances of the classification algorithm like support vector machines and artificial neural network are used for the detection of fake accounts.**

Keywords-- Fake Profile Identification, Machine Learning, Neural Network, SVM

I .INTRODUCTION

Nowadays, Instagram is dominating many aspects of social networking sites. Every day, the number of users engaging with social media changes drastically. The main advantage of Instagram is that it allows us to connect with people easily, share photos and videos, and communicate more effectively. Several online shopping stores and event planners use Instagram as a medium to gain popularity and sell their products and services. This has provided new avenues for potential attacks, such as fake identities and misinformation. A recent survey suggests that the number of accounts present on social media is much greater than the number of actual users. This indicates that fake accounts have increased in recent years. Online social media providers face difficulties in identifying these fake accounts. The need to identify these fake accounts arises because social media is flooded with bogus statistics, advertisements, etc. This paper is structured as follows: In the next section, we discuss the methodology of the proposed system. Section 3 deals with accuracy comparison and prediction, Section 4 discusses the results and analysis, and the final section covers the conclusion of the paper.

II. METHODOLOGY

The following modules are presented in this paper:

1. Information assortment
2. Uploading dataset and information preprocessing
3. Ensemble Learning strategies
4. Accuracy comparison and Prediction

2.1 Information assortment

An internet tool is employed to gather information from Instagram profiles. However, the official version of Instagram does not permit data scraping, so we created a manual dataset.

2.2 Uploading dataset and information preprocessing

Data preprocessing is the initial step to improve the model quality. The steps involved are:
Information Collection
Handling Missing Information
Handling Categorical Data
We have used Python for preprocessing.

2.3 Ensemble Learning strategies

Ensemble learning is employed to average the predictions of various models for improved prediction accuracy. This technique is particularly useful when accuracy is critical and there is no constraint on time. If we run N models, the process will be slower by a factor of N. Ensemble learning uses multiple models with different algorithms to enhance accuracy.

In our approach, we have utilized SVM and ANN for the ensemble learning process.

III. Literature Survey

S. no.	Year	Author	Focus of Paper	Key Points	Technique Used	Research Gap
1	2023	P. Durga Dr. T. Sudhakar	THE USE OF SUPERVISED MACHINE LEARNING CLASSIFIERS FOR THE DETECTION OF FAKE INSTAGRAM ACCOUNTS	Machine Learning, Instagram, Supervised Machine Learning, Spammer	K-NN, Logistic Regression, Decision Tree	Lack of deep learning machine learning models to increase accuracy of predictions.
2	2023	Fatema A. Sarhan Ebrahim Mittar	Fake Accounts Detection in Online Social Networks using Hybrid Machine Learning Models	Fake accounts, Machine Learning, SVM, ANN, Hybrid Machine Learning Models	K-NN SVM ANN Random Forest	Employ a larger dataset with more features and use of other machine learning algorithms.
3	2023	Nalia Graciella Kerrysa Ika Qutsiati Utami	Fake account detection in social media using machine learning methods	Fake accounts, social media, Machine Learning,	Various Machine Learning Techniques	Use of more larger dataset with more features in future.

IV. ACCURACY COMPARISON AND PREDICTION

The following signs indicate whether an account is real or fake:

Disproportionate Following-to-Follower Ratio: The number of accounts they follow significantly exceeds the number of their followers.

Simultaneous Postings: If multiple photos are posted on the same day, it's likely that they were posted by a bot.

Post Count: The number of posts can indicate activity levels. An account with many followers but no posts may be fake.

Inactive Account with High Follower Count: If the account has been inactive for a long period but has a high number of followers, it could be suspicious.

Lack of Personal Information: No personal information in the bio can be a red flag.

Unrealistic Posts: Lack of genuine, realistic posts may indicate a fake account.

Default Profile Picture: Having a default profile picture or no picture at all is often a sign of a fake account.

Excessive Special Characters in Bio or Username: The use of many special characters in the bio or username can also indicate a fake profile.

Type of Comments: The nature of the comments they post can also be revealing. Fake accounts often post generic or irrelevant comments.

Account Verification: Verified accounts usually belong to prominent individuals or professionals, so the lack of verification in an account claiming to be a public figure can be suspicious.

These indicators can help in identifying fake accounts and ensuring a more genuine interaction experience on social media platforms.


```
Classification Report:
              precision    recall  f1-score   support

           0       0.91      0.97      0.94        65
           1       0.96      0.88      0.92        51

    accuracy                           0.93       116
   macro avg       0.94      0.93      0.93       116
weighted avg       0.93      0.93      0.93       116

Accuracy Score: 0.9310344827586207
```

V. ALGORITHMS

The following algorithms have been used:

SVM Algorithm: The Support Vector Machine (SVM) algorithm is widely used in machine learning for predicting and classifying data. It handles both linear and non-linear problems effectively and is applicable to many real-world issues. The core idea of SVM is to create a hyperplane that separates data into different classes. In machine learning, the radial basis function (RBF) kernel is a popular choice for kernelized learning algorithms, especially in SVM classification. The RBF kernel is effective in transforming the input space into a higher-dimensional space, where it becomes easier to classify the data points.

ANN Algorithm: The Artificial Neural Network (ANN) model, inspired by the human brain, is utilized for identifying fake Instagram profiles. Trained on diverse data, the ANN learns intricate patterns within user profiles, improving its capability to distinguish between authentic and fake accounts. This model excels in capturing complex relationships, contributing to robust and accurate detection. It plays a critical role in enhancing Instagram's security by identifying deceptive profiles, particularly within the Generation Z demographic.

Ensemble Model: The Ensemble model combines the strengths of both the SVM and ANN models to improve overall predictive performance. By integrating the diverse perspectives of individual models, the Ensemble approach achieves a synergistic effect, enhancing the accuracy and robustness of the fake account detection system. This collaborative method ensures a more comprehensive evaluation, addressing the nuances of different account characteristics for a more reliable outcome.

VI. RESULT AND ANALYSIS
6.1 Dataset Description
The Instagram data taken from Kaggle website for our research.
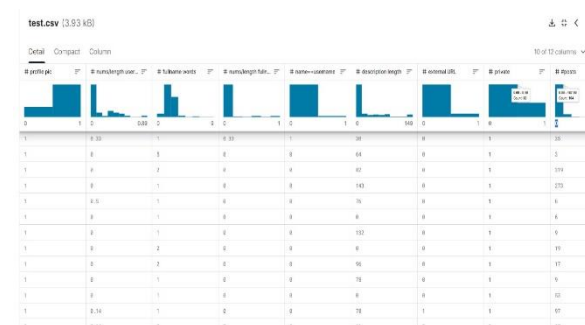

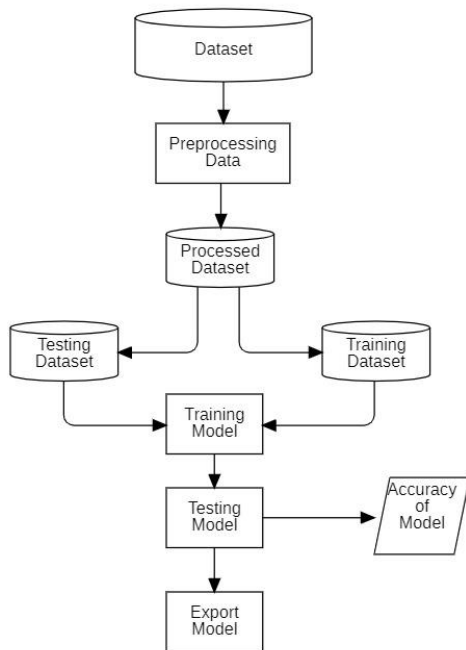
Figure 1: Instagram Training Dataset

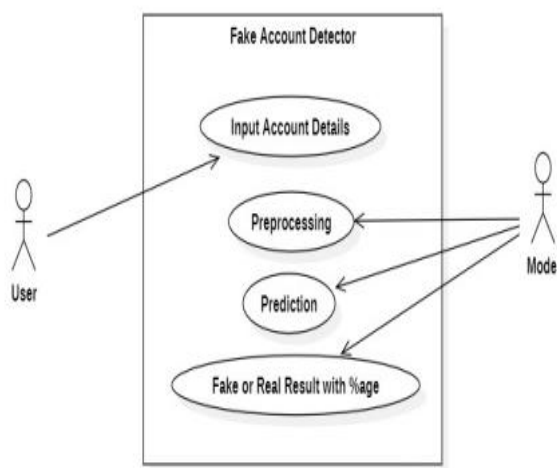Figure 2: Activity diagram for training workflow

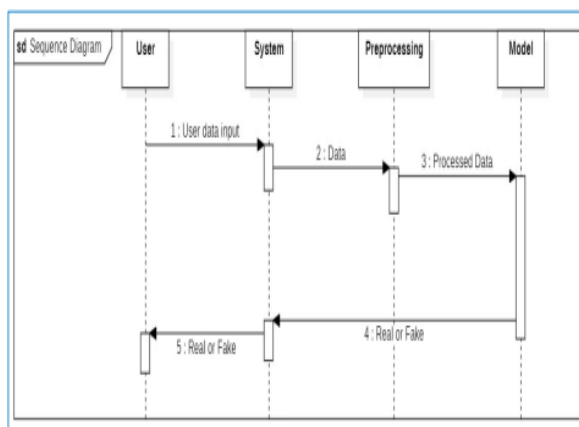
Figure 3: Use Case Diagram

Figure 4: Sequence Diagram

In this paper multiple machine learning algorithms are using against detecting fake accounts in Instagarm, among these algorithms Ensemble model performed well and it give 93.93% accuracy.

VII. CONCLUSION

Algorithms will be trained using historical data from both fake and genuine accounts. When new test data is introduced, the trained model will be applied to determine whether the new account details belong to genuine or fake users. Machine learning-based methods are employed to detect false accounts that can mislead people.

The dataset is pre-processed using various Python libraries, and a comparison form is obtained to select the most appropriate algorithm for the specified dataset. Detecting fake accounts on social media platforms is driven by a variety of machine learning algorithms. The performances of Artificial Neural Networks (ANN) and Support Vector Machines (SVM) are used for this detection.

While achieving 100% accuracy in predicting all accounts correctly is nearly impossible, this project performs well even when some features are missing, achieving an accuracy rate of 93%.

VIII. REFERENCES

[1] P. Durga, Dr. T. Sudhakar, "THE USE OF SUPERVISED MACHINE LEARNING CLASSIFIERS FOR THE DETECTION OF FAKE INSTAGRAM ACCOUNTS," 2023.

[2] Fatema A. Sarhan, Ebrahim Mittar, "Fake Accounts Detection in Online Social Networks using Hybrid Machine Learning Models," 2023.

[3] Nalia Graciella Kerrysa, Ika Qutsiati Utami, "Fake account detection in social media using machine learning methods," 2023

[4] Instagram fake genuine accounts 2019, Accessed on 15 October 2023, https://www.kaggle.com/datasets/sizlingdhairya1/instagram-fake-genuine-accounts

[5] C. Akyon and M. Esat Kalfaoglu, "Instagram Fake and Automated Account Detection," 2019 Innovations in Intelligent Systems and Applications Conference (ASYU), 2019

[6] B. Erˌsahin, Ö. Aktaˌs, D. Kılınç, and C. Akyol, "Twitter fake account detection," in 2017 International Conference on Computer Science and Engineering (UBMK). IEEE, 2017

[7] F. Benabbou, H. Boukhouima, and N. Sael, "Fake accounts detection system based on bidirectional gated recurrent unit neural network," International Journal of Electrical and Computer Engineering (IJECE), 2022

[8] V. Sarala and G. Sandhya, "Spammer Detection and fake user Identification on Social Networks," Journal of Engineering Sciences, vol. 13, no. 8, 2022.

[9] S. Sheikhi, "An Efficient Method for Detection of Fake Accounts on the Instagram Platform," Revue d'Intelligence Artificielle, vol. 34, no. 4, 2020

[10] D. B. S. Reddy, D. R. A. Nayana, G. Sahaja, and G. S. Priya, "Artificial Intelligence Tool for Fake Account Detection from Online Social Networks," Turkish Journal of Computer and Mathematics Education (TURCOMAT), vol. 14, no. 1, 2023