Fake Profile Detection on social media Using Python

[1] G.Sravanthi, [2] V. VENKATA SAHITHI, [3] V. YASWANTH, [4] P. SINDHUJA, [5] T. AKHILA

[1] Assistant Professor, Department of ECE, Sri Indu College of Engineering & Technology, Hyderabad.

[2,3,4,5] U.G. Scholar, Department of ECE, Sri Indu College of Engineering & Technology, Hyderabad.

ABSTRACT:

Online social networks represent one of the largest digital ecosystems, with more than 4 billion users worldwide. With this rapid expansion, the risk of maintaining data integrity and security has significantly increased. Social media platforms face numerous security challenges, as users often provide limited personal information, making identity verification difficult. This creates opportunities for scammers to generate fake accounts and misuse them to steal sensitive information, spread misinformation, or harm other users.

As a result, detecting fake profiles in social networks has become a complex and important task. This project utilizes various machine learning and deep learning techniques applied to different datasets to identify the most effective model for detection. Additionally, a web-based system is developed to determine whether a given user profile is real or fake.

The system is designed with four modules representing major social media platforms that are rapidly gaining users. Each module contains a trained model specifically developed for detecting fake profiles on that platform. Currently, the system has been implemented for Instagram, enabling accurate identification of fake profiles through the developed model with improved performance.

Keywords : Fake profile, Instagram , Machine Learning , Random Forest Classifier , Prediction.

1. Introduction

Our project aims to detect the fake profiles of various famous social media applications like Instagram, Facebook, Twitter and Linked In. By using different machine learning and deep learning techniques on different datasets to find the best model and making an website to detect whether the user given profile real or fake. The website consists of 4 modules which are 4 major social networks which are gaining users rapidly. Each module consisting a model which is trained to detect whether the profile is real or fake.

2. Literature Review

Instagram is a free social platform for the user interaction across the platform. It is similar to other social media platforms, where the people who create an account will have a profile and feed where they can share photos and videos through that.[4] The results were obtained with the minimum available profile data. In addition, in comparison with the other methods that use the same amount and kind of data, the proposed deep neural network gives an increase in accuracy of roughly 14%.[3]Detecting such profiles is essential in terms of security. There are very few researches have been done to identify fake profiles on social media platforms. Various machine learning (ML) methods are used to do this task[2]

Facebook is considered the biggest online social network. Facebook has 2.2 billion monthly active users as of the first month of 2021, whereas Twitter has 330 million monthly active users and 145 million daily active users.[1]Twitter allows a user to post 280 characters to express their feelings and thoughts. Therefore, malicious users use shortened links and shortened URLs and they also express their ideas with repetitive words to grab the real users' attentions. Fake users have their fake patterns and can be easily checked them by carefully examining their tweeting patterns[5]LinkedIn has the third-highest visit rate, with more than 200 million visitors per month, and serves as the world's largest professional network.[6]

3. Dataset:

4. Methodology

Using Machine Learning to predict the fake profiles where the dataset has been trained with different models like SVM , logistic Regression and Random Forest Classsifier . In which Random Forest Classifier has given an accuracy of 98%.So.Random Forest Classifier is selected for the real time prediction.

4.1 Random Forest Classifier

Random Forest is a machine learning algorithm which is a supervised learning technique. It is used for both Classification and Regression problems in Machine Learning. It is based on the concept of ensemble learning, where it is a process of combining multiple classifiers to solve a problem and to improve the performance of the model.

It is a classifier that contains a number of decision trees on the given dataset and takes the average to improve the accuracy of that dataset. Instead of depending on one decision tree for the dataset to predict the random forest classifier takes the prediction from each tree and based on the majority votes of decision trees across the dataset, it predicts the final output.The more no. of trees in the forest leads to high accuracy and prevents overfitting.
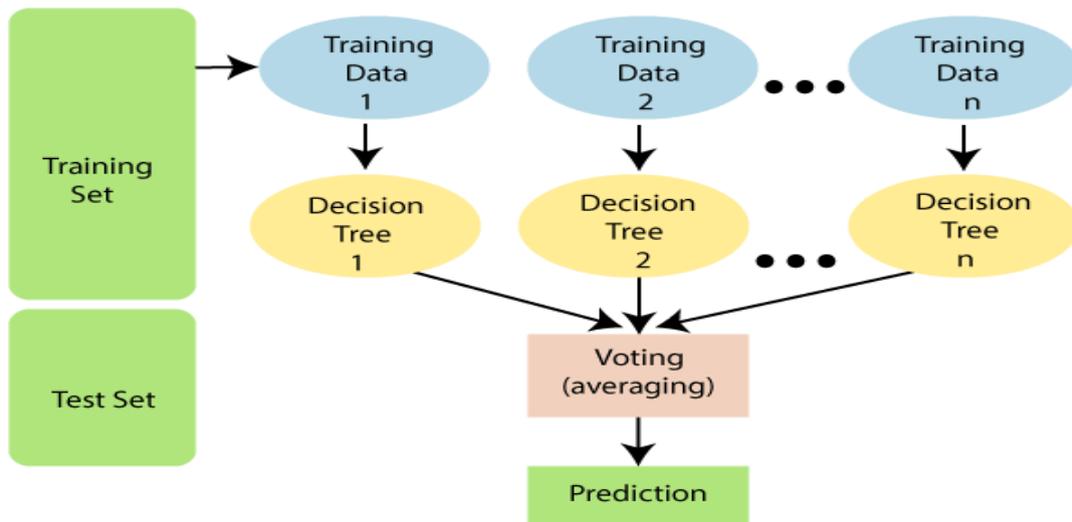

Fig 1. Random Forest Classifier

This classifier takes the input as Post Counts, Following Count, Follows Count, Verified and private and the random forest takes the prediction from each tree across the data and predicts the output.

4.2 Training

The training of the model includes separating the dataset in the ratio of 80:20 , 80 for the training dataset and 20 for the testing dataset. Then those independent and dependent features will be fitting into the model and the model is trained by the training dataset.

4.3 Testing

After the model is been trained then the model can be tested with the remaining test dataset with predict function .Then by using metrics the accuracy is found which is 98%.

4.4 Real Time account prediction

By using API(Application Programming interface) which is APIFY Instagram-profile-scarpper the client is intergrated in the python code to access and implement real time account prediction.
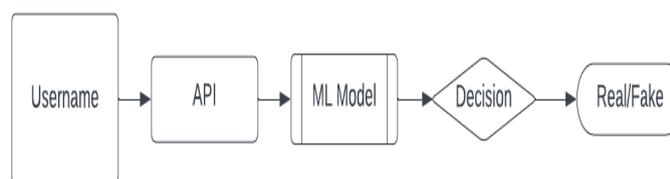

Fig 2 . Flow of data to the model

4.5 Deployment using Streamlit

The code is deployed using streamlit for frontend to provide user interface to predict the profile is real or fake.
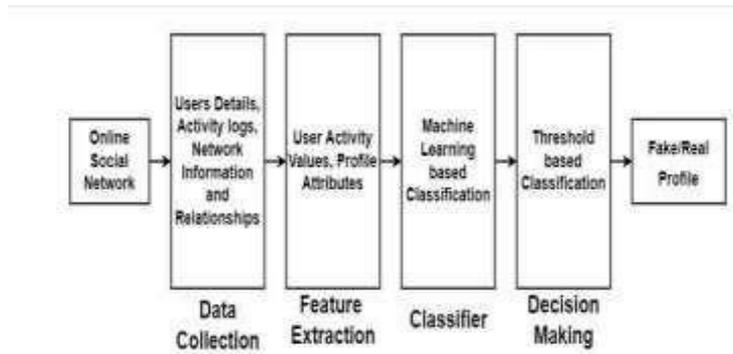
4.6 Architecture

Fig 3. Architecture

5. Result

5.1 Output Screens

Fig 4. Frontend for Prediction

Fig 5 . Predicting a Real Account by username

Fig 6. Predicting a Fake Account by username

6. Conclusion

Fake accounts are dangerous for social platforms since they may alter concepts like popularity and influence on people .By training the model with various algorithms and large datasets helps and maintains accuracy at real time implementation .The proposed methodology has the potential to be of great assistance in the fight against fraud on social networking sites.

7. Future Work

By suitable API's for the other modules which are twitter , Facebook and LinkedIn can also be developed to detect the fake profiles across the platforms. Those module implementations are the next part of our project to enhance the platform usage by detecting the fake accounts.

8. References:

[1] Ahmad Homsi, Joyce Al Nemri, Nisma Naimat, Hamzeh Abdul Kareem. Detecting Twitter Fake Accounts using Machine Learning and Data Reduction Techniques

[2] Bharti, Nasib Singh Gill, Preeti Gulia. Exploring machine learning techniques for fake profile detection in online social networks

[3] Daniyal Amankeldin , Lyailya Kurmangaziyeva , Ayman Mailybayeva. Deep Neural Network for Detecting Fake Profiles in Social Networks

[4] Er. Pranay Meshram, Rutika Bhambulkar, Puja Pokale . Automatic Detection of Fake Profile Using Machine Learning on Instagram

[5] Myo Swe . Fake Accounts Detection on Twitter Using Blacklist

[6] Priyanka Kondeti, Lakshmi Pranathi Yerramreddy , Anita Pradhan and Gandharba Swain. Fake Account Detection using Machine Learning

[7] Saeid Sheikhi . An Efficient Method for Detection of Fake Accounts on the Instagram Platform

[8] Shalinda Adikari. Identifying Fake Profiles in LinkedIn