Detection of Fake and Clone Accounts on Twitter

Kavitha S
Computer Science
(of Affiliation)
K S Institute of Technology
(of Affiliation)
Bangalore, India

Sharanya H
Computer Science
(of Affiliation)
K S Institute of Technology
(of Affiliation)
Bangalore, India

Dr. Deepa S R
Computer Science
(of Affiliation)
K S Institute of Technology
(of Affliliation)
Bangalore, India

Rajashree Shivakumar
Computer *Science*
(of Affiliation)
K S Institute of Technology
(of Affiliation)
Bangalore,India

Kruthika B M
Computer Science
(of Affiliation)
K S Institute of Technology
(of Affiliation)
Bangalore,India

***Abstract*— Online Social Network (OSN) has become an important aspect of many people's lives in recent years. Online social networks are used for a variety of purposes, including communication, promotion, advertising, news, and agenda development. However, some illegitimate identities on social networking sites are posing a security risk to legitimate users. Twitter is one such social networking service that is popular among users, but it also contains a large number of fake and clone accounts that pose a threat to genuine users. In this paper, a detection system has been proposed which detects the existing fake accounts on Twitter using a supervised machine learning algorithm and detects the duplicate accounts while they are being created, as clones and prevents the creation of clone accounts.**

***Keywords*— *Online Social Network, Fake, Clone, Supervised Machine Learning Algorithm*.**

I. INTRODUCTION

Social networking phenomenon has grown extremely since the last twenty years. During this period of growth, online social networks have developed a plethora of online activities that appeal to a wide range of people. However, they are also affected by the rise in the number of false and clone account.

Online Social Network (OSN) users share a lot of career and personal information in the network. If this information goes to the hands of attackers, the after effects can be dangerous. A majority of online social network users are unaware of the safety threat that prevails in social media and are thus easily targeted by these attackers. Twitter is one such social networking service that is popular among users, but it also contains a large number of fake and clone accounts that pose a threat to the genuine users.

The account created in the name of a person or a company that does not exist on social media in order to carry out destructive acts is known as fake account. In fake accounts people usually lie about their age, gender and also use profile pictures taken from the internet in order to hide their identity. They act in a prohibited manner such as attempting to deceive or mislead people by posting harmful links and aggressive following behaviors can also be found in such accounts like mass following etc.

Clone account is a duplicate account of existing user, created by using the original user's details like, name, profile picture, date of birth, place, etc. These kind of accounts are created to harm the identity of the original users.

Due to the extreme importance of the impact of social media on the society, this paper aims at detecting the fake and clone accounts on Twitter to prevent the problems caused by these accounts to the genuine users.

II. LITERATURE SURVEY

Nowadays, Fake and Clone accounts have become a very serious issue on social media. So, a detection method is very much necessary in order to overcome the problems caused by the frauds that create these fake and clone accounts which is a threat to the genuine users. Many authors have worked in this area and have proposed various methods to identify these types of profiles in social networks. Some of these methods are discussed below.

Two unique approaches for detecting cloned profiles have been proposed by Piotr Brodka, Mateusz Sobas, and Henric Johnson. The first method is based on attribute value similarity between original and cloned profiles, while the second method is based on network relationships. A victim will be selected from those who believe their profile has been copied. Then treating name as primary key, a search is made for profiles with the same name as that of victim, using query search. Potential clone (Pc) and the Victim profile (Pv) are compared and similarity S is calculated. If S(Pc, Pv) > Threshold, then profile is suspected to be a clone. In the verification step, the user does it manually as

he knows which one is the original profile and which is the duplicate one.

Supraja Gurajala, Joshua S. White, Brian Hudson and Jeanna N. Matthews, have proposed a system: A crawler was used to collect and analyse a big Twitter user profile database of 62 million accounts in order to better understand the features of false account creation. User accounts were grouped based on matched multiple-profile-attributes, patterns in their screen names, and an update-time distribution filter, resulting in a highly trustworthy fake profile collection. A subset of the accounts identified as fake by the algorithm were manually inspected and verified as all being fake accounts (based on their Tweet activity).

Ahmed El Azab, Amira M Idrees, Mahmoud A Mahmoud, Hesham Hefny, have proposed a classification method for detecting fake accounts on Twitter. In the initial stage, they gathered several useful features for the detection process from various studies and filtered and weighted them. Various studies are carried out in order to find the smallest collection of features that produce accurate results. Only seven variables were chosen from a total of 22 to efficiently detect bogus accounts, and these factors were used in a classification technique. A comparison of the classification techniques based on results is made and the one which provides most accurate result is selected.

In the suggested study, Arpitha D, Shrilakshmi Prasad, Prakruthi S, and Raghuram A S, have used multiple algorithms for different parameters such as name, age, address, and email. Duplicate profiles can simply be recognised using these methods. All of the algorithms increase the likelihood of finding the correct match.

Sowmya P and Madhumita Chatterjee, have suggested a method in which a set of rules were employed for fake detection, which when implemented can categorize fake and authentic profiles. Clone identification was carried out using Similarity Measures and the C4.5 method, and the results were compared. Clone identification using Similarity Measures outperformed C4.5, detecting the majority of the clones supplied into the system. In this work they have considered only the profile attributes for fake and clone detection.

III. PROPOSED METHODOLOGY

A. Fake Account Detection:

Firstly, all the libraries required for performing data preprocessing are imported, and then the data is read from the data set. The attributes and values of the dataset are analyzed. After analyzing the data set, handling of categorical features is done by elimination of few features, which are not of much importance.

The handling of missing data in the dataset is done by eliminating the rows which had missing values. Then, all the data values are concatenated into a single data frame. The dataset is split into train data and test data, where train data consists of 80% of the dataset and test data is 20%.

Naïve Bayes Classifier is imported and the train data is fit into the algorithm. The values from test dataset is passed to the Naïve Bayes Classifier, function predict, which will be predicting an account to be fake or real.

A GUI window is created where the values of the considered account attributes are given as inputs and is predicted whether it is Fake or Real. Tkinter is imported which is a graphical user interface for python, using which windows, buttons, show texts etc. can be created. Labels are created for each attribute where the values of the attributes can be entered. Then, in the predict function, it gets all the values from the window uses Naïve Byes algorithm to predict whether the account is Fake or Real and then, it is displayed on the window.

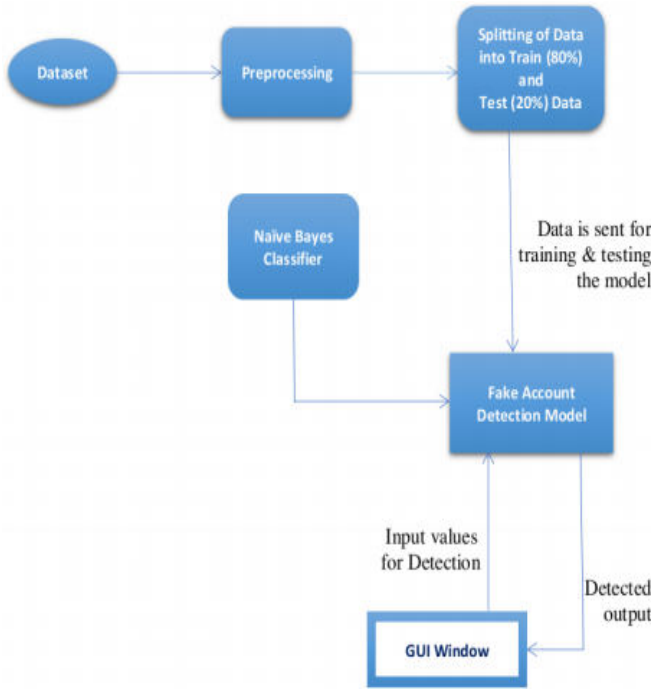

Fig 1: Fake Account System Architecture

B. Evaluation Metrics:

The performance evaluation of our fake account detection model is done using confusion matrix.

TOTAL NUMBER OF ACCOUNTS CHECKED	***1276***
No. of fake accounts detected as fake	1056
No. of real accounts detected as fake	12

No. of fake accounts detected as real	5
No. of real accounts detected as real	203

Table 1: Performance Evaluation of Fake Account Detection Model

C. Clone Account Detection:

A detection system is created in form of a webpage to detect and prevent the clone accounts from being created. Here, python is used for programming, HTML and sqlite3 is used for making simple login for preventing the creation of clone account.

SQLite3 is imported and a database is created and within the database a user table is created with user credentials such as name, email, DOB and password. Then, an HTML document named user register is created. User register is created for users to register. It only takes the values if the user details such as name and email id is not of existing user. If username or email id given by the new user already exists in the database then, it is detected as clone account and the same is displayed and it does not allow the account to be created with the same user information.

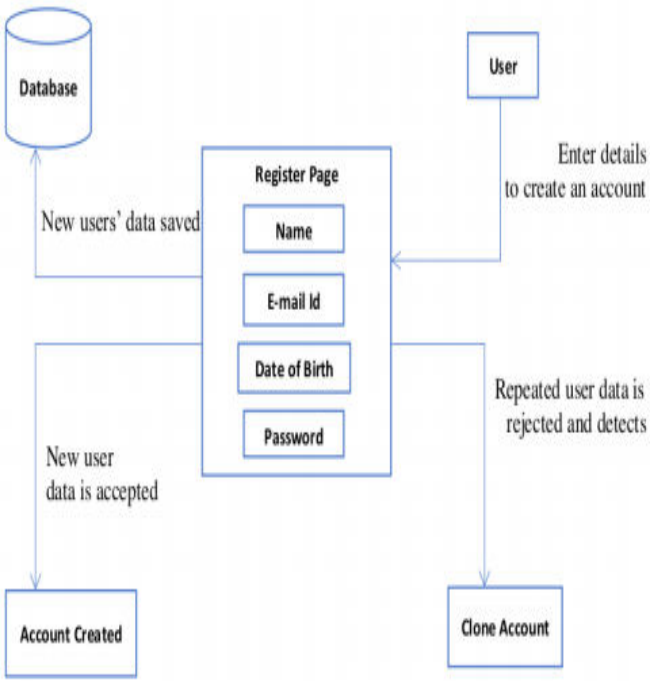

Fig 2: Clone Account System Architecture

IV. CONCLUSION

Fake and clone accounts have become a serious threat to social media users. So, here a detection method is proposed which detects whether the twitter account is fake or real. Fake accounts detection is carried out using Naïve Bayes Classifier by considering a few profile attributes like, status count, following count, likes count and listed count. Clone accounts are detected and are prevented from being created if the user name and the email id are of already existing users. In future, fake account detection work can be enhanced by taking tweets and profile pictures as well into consideration. For clone account detection datasets can be collected and the existing clone accounts on twitter can be detected.

V. ACKNOWLEDGEMENT

We would like to express our heartfelt gratitude to Dr. Deepa S R, for her invaluable guidance, insightful comments, helpful information and suggestions, hence improving our knowledge.

VI. REFERENCES

[1] Piotr Bródka, Mateusz Sobas and Henric Johnson, "Profile Cloning Detection in Social Networks", 2014 European Network Intelligence Conference.

[2] Supraja Gurajala, Joshua S. White, Brian Hudson and Jeanna N. Matthews "Fake twitter accounts: profile characteristics obtained using an activity-based pattern detection approach", International Conference on Social Media & Society, ACM 2015.

[3] Ahmed El Azab, Amira M Idrees, Mahmoud A Mahmoud, Hesham Hefny, "Fake Account Detection in Twitter Based on Minimum Weighted Feature set", World Academy of Science, Engineering and Technology, International Journal of Computer and Information Engineering Vol:10, 2016.

[4] Arpitha D, Shrilakshmi Prasad, Prakruthi S, Raghuram A S, "Python based machine learning for profile matching", International Research Journal of Engineering and Technology (IRJET), Volume: 05 Issue: 03, 2018.

[5] Sowmya P and Madhumita Chatterjee, "Detection of Fake and Cloned Profiles in Online Social Networks", Proceedings 2019: Conference on Technologies for Future Cities (CTFC).