# Fake User Detection In Social Network

**Subhratara Sahoo[1], Saraswati Patkar[2], Pranjali Kushekar[3]**

*[1-3]Dept. of Computer Engineering, PHCET College, Maharashtra, India*

---------------------------------------------------------------------***---------------------------------------------------------------------

**Abstract -** *Social networking sites engage millions of users around the world. The users' interactions with these social sites, such as Twitter and Facebook have a tremendous impact and occasionally undesirable effect for daily life. Twitter, for example, has become one of the most excessive used platforms of all times and therefore allows an unreasonable amount of spam. Fake users send undesired tweets to users to promote services or websites that not only affect legitimate users but also interrupt resource consumption. Moreover, the possibility of expanding invalid information to users through fake identities has increased that results in the unrolling of harmful content. Recently, the detection of spammers and identification of fake users on Twitter has become a common area of research in contemporary online social Networks (OSNs).The Application Domain of the following project was Community Detection. Community detection is key to understanding the structure of complex networks, and ultimately extracting useful information from them. In this project, we came up with a framework through which we can detect a fake profile using machine learning algorithms so that the social life of people become secured.*

***Key Words*: Machine Learning, OSNs(Online social Networks), Community Detection, spammers, legitimate.**

## 1.INTRODUCTION

A social networking site is a website where each user has a profile and can keep in contact with friends, share their updates, meet new people who have the same interests. Social networking sites are growing rapidly and changing the way people keep in contact with each other. The online communities bring people with same interests together which makes users easier to make new friends In the present generation, everyone in society has become associated with the Online Social Networks(OSN). These OSN have made a drastic change in the way we pursue our social life. Making new friends, keeping in contact with them and knowing their updates has become easier. But with the rapid growth of social media many problems like fake profiles, online impersonation have also grown. There are no feasible solution existing to control these problems. Fake accounts can be either humangenerated, computer generated( also referred as "bots"), or cyborgs. A cyborg is half-human, half-bot account. Such an account is manually opened by a human, but from then onwards the actions are automated by a bot. To become member of the OSN the user has to create his profile by entering information like name, photo, date of birth, Email ID, graduation details, place of work, home town, interests and so on. Some of the fields are mandatory and some are optional and it varies from one OSN to the another. These websites are popular because of people's interest in finding friends, sharing pictures, tagging people in group photos, sharing their ideas and views on common topics, maintain good business relationship and general interest with others. In this project we came up with a framework in which automatic detection of fake profiles is possible and is efficient.

### 1.3 SOCIAL IMPACT

The increased demand of social sites permits users to collect abundant amount of information and data about users. Huge volumes of data available on these sites also draw the attention of fake users. With the evolution of OSNs, the need to study and analyze users' behaviors in online social platforms has intensified. Many people who do not have much information regarding the OSNs can easily be tricked by the fraudsters. There is also a demand to combat and place a control on the people who use OSNs only for advertisements and thus spam other people's accounts.

### 1.2 ISSUES

The social networking sites are making our social lives better but nevertheless there are a lot of issues with using these social networking sites. The issues are privacy, online bullying, potential for misuse, trolling, etc. These are done mostly by using fake profiles.

### 1.3 OBJECTIVE

In this project, we came up with a framework through which we can detect a fake profile using machine learning algorithms so that the social life of people become secured.

## 2. LITERATURE SURVEY

Fake profiles are the profiles which are not genuine i.e. they are profiles of persons who claim to be someone they are not, doing some malicious and undesirable activity, causing problems to the social network and fellow users. Why do people create fake profiles ?

- Social Engineering.
- Online impersonation to defame a person.
- Advertising and campaigning a person.

## 3. PROPOSED SYSTEM

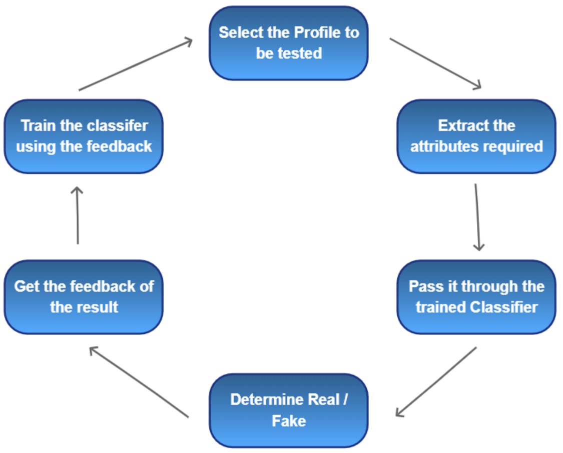

Select the Profile to be tested
Extract the attributes required
Pass it through the trained Classifier
Determine Real / Fake
Get the feedback of the result
Train the classifer using the feedback

**Fig-3**: Framework for detection of fake profiles

1. Classification starts from the selection of profile that needs to be classified.

2. Once the profile is selected, the useful features are extracted for the purpose of classification.

3. The extracted features are then fed to trained classifier.
4. Classifier is trained regularly as new data is fed into the classifier.

5. Classifier then determines whether the profile is genuine or fake.

6. The result of classification algorithm isthen verified and feedback is fed back into the classifier.

7. As the number of training data increases the classifier becomes more and more accurate in predicting the fake profiles.

## 4. DETECTION STRATEGY

### A. Web Scraper

Web Scraper is used to extract data from a website. We extract data such as login activity, Total Likes, Total Comments, Number of posts, Number of followings and Number of Followers.

### B. Calculation of Engagement rate

An engagement rate is a metric that measures the level of engagement of a Post or Story received on social media. It is the percentage by which the audience interact with a post. By checking the number of interactions with the number of followers we can evaluate the engagement rate. Interactions can be of likes, comments, and shares. Most Fake accounts will boast of 1000s of followers and a very minimum number of likes. Since the engagement rate is relatively calculated, comparisons between popular accounts and semipopular accounts are comparatively easy.

$$\text{Engagement Rate Percentage} = \left[\frac{\text{Total number of Interaction}}{\text{Total number of Followers}}\right] *100$$

### C. Artificial Activity

Normal social media activities such as liking, commenting and sharing turns into an artificial activity when the frequency of the above mentioned are very high. Around the clock activity also signifies that the account is used by a Bot. At this stage, we look into the number of likes, comments, and shares this particular account has made since its creation. If an enormous amount of likes or comments are found, then that account will be considered as fake.

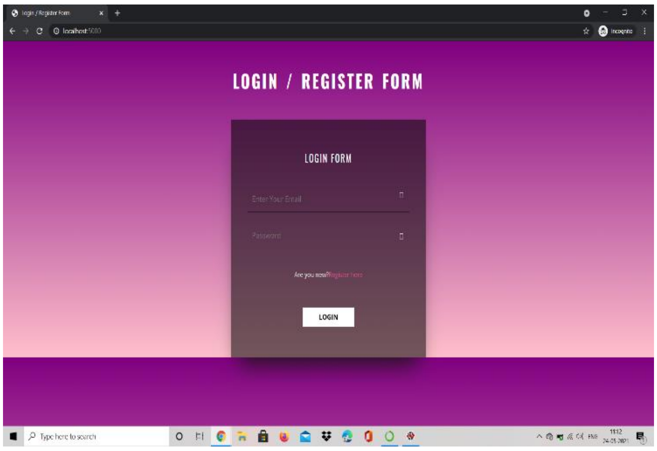
### D. Spam Comments

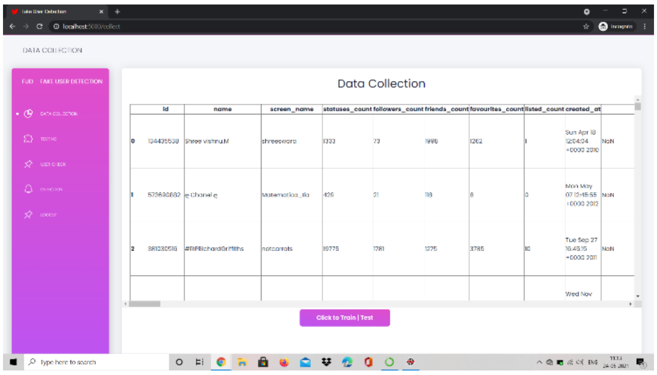
BOT comments are always known to be very Generic and often lack Substance. At this stage, the comments made from the account will be gone through in a detailed manner. Total number of comments by the user made since the creation of the account will be compared with average comments of users in that particular OSN's. If there is a big difference the account may be considered fake. Commenting links will lead to the account being termed as Fake account. Same or Similar type of comments will also be considered as spam comments.

### E. Detection of Fake Accounts

In this step, we combine all the data we extracted from the website. In this paper we mainly focus on engagement rate, artificial activity and spam comments. The data collected using web scraper is used to compute the values for the factors mentioned above. Using these factors different decision tress is formed. Using the algorithm and with the formed decision trees fake accounts are detected.

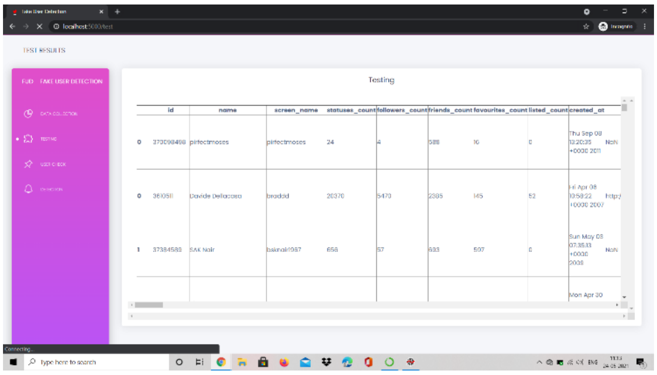
## 5. RESULT





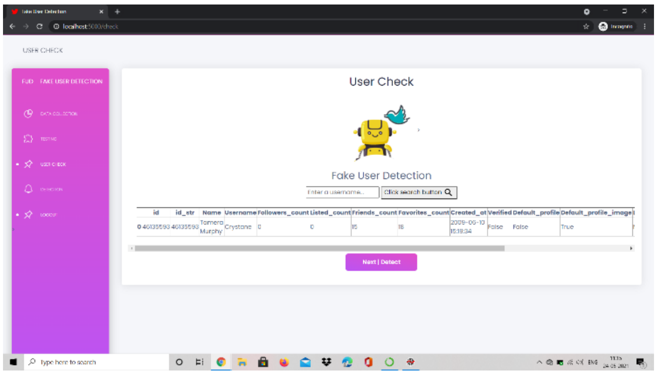
Data Collection

| | id | name | screen_name | statuses_count | followers_count | friends_count | favourites_count | listed_count | created_at | |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 134435538 | Shree vishnu.M | shreeswara | 1333 | 73 | 1998 | 1262 | 1 | Sun Apr 18 12:04:04 +0000 2010 | NaN |
| 1 | 573690682 | ღ Chanel ღ | Matematica_tia | 429 | 21 | 118 | 8 | 0 | Mon May 07 12:45:55 +0000 2012 | NaN |
| 2 | 381030516 | #RIPRichardGriffiths | notcarrots | 19775 | 1781 | 1275 | 3785 | 10 | Tue Sep 27 16:45:15 +0000 2011 | NaN |
| | | | | | | | | | Wed Nov | |



Testing

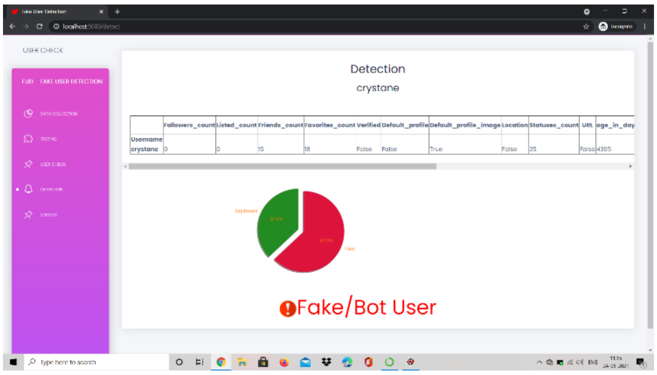
| | id | name | screen_name | statuses_count | followers_count | friends_count | favourites_count | listed_count | created_at | |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 370098498 | pirfectmoses | pirfectmoses | 24 | 4 | 588 | 16 | 0 | Thu Sep 08 13:20:35 +0000 2011 | NaN |
| 0 | 3610511 | Davide Dellacasa | braddd | 20370 | 5470 | 2385 | 145 | 52 | Fri Apr 08 10:58:22 +0000 2007 | http:/ |
| 1 | 37384589 | SAK Nair | bsknair1967 | 656 | 57 | 633 | 507 | 0 | Sun May 03 07:35:13 +0000 2009 | NaN |
| | | | | | | | | | Mon Apr 30 | |



User Check

Fake User Detection

| | id | id_str | Name | Username | Followers_count | Listed_count | Friends_count | Favorites_count | Created_at | Verified | Default_profile | Default_profile_image |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 40135593 | 40135593 | Tamera Murphy | Crystane | 0 | 0 | 15 | 18 | 2009-05-10 15:19:34 | False | False | True |



Detection

crystane

| | Followers_count | Listed_count | Friends_count | Favorites_count | Verified | Default_profile | Default_profile_image | Location | Statuses_count | URL | age_in_day |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Username | | | | | | | | | | | |
| crystane | 0 | 0 | 15 | 18 | False | False | True | False | 25 | False | 4365 |

Fake/Bot User

## 6. CONCLUSION

In this paper, we performed a review of techniques used for detecting spammers on Social Networks. In addition, we also presented a taxonomy of spam detection approaches and categorized them as fake content detection, URL based spam detection, spam detection in trending topics, and fake user detection techniques. We also compared the presented techniques based on several features, such as user features, content features, graph features, structure features, and time features.

## 7. REFERENCES

1. Andrew Ng, “Machine Learning” https://www.coursera.org/learn/machine-learning

2. freelancer.com

3. “Support Vector Machine”

https://en.wikipedia.org/wiki/Support_vector_machine

4. “Chapter-2 Support Vector Machine(SVM) theory” https://medium.com/machine-learning-101/chapter-2-svm-support-vector-machin e-theory-f0812effc72

5. Shahzeb Haidar “Facebook Immune System: A summary”
http://home.iitk.ac.in/~shaidar/cs300/4B/4B.pdf

6. Jason Brownlee ”Support Vector Machines for Machine Learning” https://machinelearningmastery.com/support-vector-machines-for-machine-learni ng/

7. “Social Bot”

https://www.techopedia.com/definition/27811/socialbot

8. “Convolutional Neural Networks”

http://cs231n.github.io/convolutional-networks/