IJARETY

# International Journal of Advanced Research in Education and TechnologY (IJARETY)

Volume 11, Issue 6, November-December 2024

Impact Factor: 7.394

INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

INNO SPACE
SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर
NISCAIR

DOI:10.15680/IJARETY.2024.1106023

# Fake User Identification on Social Network

**Dr.Atul Kumar Ramotra[1], K.Rithvika[2], M.Krithika[3], I.Vishnu Vignesh[4]**

Assistant Professor, Department of CSE, Guru Nanak Institute of Technology, Hyderabad, India[1]

Student, Department of CSE, Guru Nanak Institute of Technology, Hyderabad, India[2,3,4]

**ABSTRACT:** Online Social Network (OSN) is a network hub where people with similar interests or real world relationships interact. As the popularity of OSN is increasing, the security and privacy issues related to it are also rising. Fake and Clone profiles are creating dangerous security problems to social network users. Cloning of user profiles is one serious threat, where already existing user's details are stolen to create duplicate profiles and then it is misused for damaging the identity of original profile owner. They can even launch threats like phishing, stalking, spamming etc. Fake profile is the creation of profile in the name of a person or a company which does not really exist in social media, to carry out malicious activities. In this paper, a detection method has been proposed which can detect Fake and Clone profiles in Twitter. Fake profiles are detected based on set of rules that can effectively classify fake and genuine profiles. For Profile Cloning detection two methods are used. One using Similarity Measures and the other using C4.5 decision tree algorithm. In Similarity Measures, two types of similarities are considered – Similarity of Attributes and Similarity of Network relationships. C4.5 detects clones by building decision tree by taking information gain into consideration. A comparison is made to check how well these two methods help in detecting clone profiles.

**KEYWORDS:** Online Social Network (OSN), Fake Profiles, Clone Profile Detection, Similarity Measures, C4.5 Decision Tree Algorithm

## I. INTRODUCTION

ONLINE Social Networks (OSN) like Facebook, Twitter, LinkedIn, Instagram etc are used by billions of users all around the world to build network connections. The ease and accessibility of social networks have created a new era of networking. OSN users share a lot of information in the network like photos, videos, school name, college name, phone numbers, email address, home address, family relations, bank details, career details etc. This information if put into hands of attackers, the after effects are very severe. Most of the OSN users are unaware of the security threats that exist in Sowmya P is with the Department of Computer Engineering, Pillai College of Engineering, University of Mumbai, Maharashtra, India (email: sowmya@mes.ac.in). Madhumita Chatterjee is with the Department of Computer Engineering, Pillai HOC College of Engineering and Technology, University of Mumbai, Maharashtra, India (e-mail: mchatterjeee@mes.ac.in). the social networks and easily fall prey to these attacks. The risks are more dangerous if the victims are children. In Profile Cloning attack, the profile information of existing users are stolen to create duplicate profiles and these profiles are misused for spoiling the identity of original profile owners [1- 6]. There are two types of Profile Cloning namely - Same Site and Cross Site Profile Cloning [1,7-9].

If user credentials are taken from one Network to create a clone profile in same Network then it is called Same Site profile cloning [1,10-12]. In Cross Site profile cloning, attacker takes the user information from one Network to create a duplicate profile in other Network in which the user is not having any account [1,13-15].

As the registration process in social networks have become very simple in order to attract more and more users, the creation of fake profiles are also increasing in an alarming rate. An attacker creates a fake profile in order to connect to a victim to cause malicious activities. And also to spread fake news and spam messages.

The popularity and ease of use of social networking services have excited institutions with their potential in a variety of areas. However effective use of social networking services poses a number of challenges for institutions including long-term sustainability of the services; user concerns over use of social tools in a work or study context; a variety of technical issues and legal issues such as copyright, privacy, accessibility; etc.

Institutions would be advised to consider carefully the implications before promoting significant use of such services. Computer security (Also known as cyber security or IT Security) is information security as applied to computers and networks. The field covers all the processes and mechanisms by which computer-based equipment, information and services are protected from unintended or unauthorized access, change or destruction. Computer security also includes protection from unplanned events and natural disasters. Otherwise, in the computer industry, the term security -- or the phrase computer security -- refers to techniques for ensuring that data stored in a computer cannot be read or compromised by any individuals without authorization. Most computer security measures involve data encryption and passwords. Data encryption is the translation of data into a form that is unintelligible without a deciphering mechanism. A password is a secret word or phrase that gives a user access to a particular program or system.

Fake and clone profiles have become a very serious social threat. As information like phone number, email id, school or college name, company name, location etc are readily exposed in social networks, hackers can easily hack this information to create fake or clone profiles. They then try to cause various attacks like phishing, spamming, cyberbullying etc. They even try to defame the legitimate owner or the organisation. So, a detection method has been proposed which can detect both fake and clone profiles in order to make the social life of the users more secure.

If user credentials are taken from one Network to create a clone profile in same Network then it is called Same Site profile cloning. In Cross Site profile cloning, attacker takes the user information from one Network to create a duplicate profile in other Network in which the user is not having any account. As the registration process in social networks have become very simple in order to attract more and more users, the creation of fake profiles are also increasing in an alarming rate. An attacker creates a fake profile in order to connect to a victim to cause malicious activities. And also to spread fake news and spam messages.

## II. LITERATURE SURVEY

In their 2019 paper, Y. Xiao, D. Chen, S. Wei, Q. Li, H. Wang, and M. Xu discuseed In the online social network, the spreading process of rumor contains complex dynamics. The traditional research of the rumor propagation mainly studies the spreading process of rumor from the perspectives of rumor and participating user. The symbiosis and confrontation of rumor and anti-rumor information and the dynamic changes of the influence of anti-rumor information are not emphasized. At the same time, people’s profitability and herd psychology are also ignored. In view of the above problems, we fully consider the anti-rumor information and user’s psychological factors, construct a rumor propagation dynamics model based on evolutionary game and anti-rumor information, and provide a theoretical basis for studying the inherent laws in the spreading process of rumor. First of all, we analyze the interaction pattern and characteristic of rumor in social network. In allusion to the symbiosis of rumor and anti-rumor information and the dynamic changes of the influence of anti-rumor information, we constructed the SKIR rumor propagation model based on the SIR model.. At the same time, we combine the behavior factors and external factors of the user to build the influence of information by multivariate linear regression method, which provides the theoretical basis for the driving force of information. Finally, combining the SKIR model proposed in this paper, we get a rumor propagation dynamics model based on evolutionary game and anti-rumor information. We have proved by experiments that the model can effectively describe the propagating situation of rumor and the dynamic change rule of the influence of anti-rumor information. On the other hand, it can also reflect the influence of people’s psychology on rumor propagation.

In their 2018 paper, S. Sommariva, C. Vamos, A. Mantzarlis, L. U.-L. Dào, and D. Martinez Tyson discuseed The importance of social networking sites (SNSs) as platforms to engage in the correction of “fake news” has been documented widely. More evidence is needed to understand the popularity of health-related rumors and how Health Educators can optimize their use of SNSs. Purpose: The purpose of this study was to explore the spread of health rumors and verified information on SNSs using the Zika virus as a case study. Methods: A content analysis of Zika-related news stories on SNSs between February 2016 and January 2017 was conducted to verify accuracy (phase 1). Phase 1 was followed by an analysis of volume of shares (phase 2) and a thematic analysis of headlines (phase 3). Results: Rumors had three times more shares than verified stories. Translation to Health Education Practice: Misinformation on SNSs can hinder disease prevention efforts. This study shows how information circulating on SNSs can be analyzed from a quantitative and qualitative standpoint to help Health Educators maximize the use of online communication platforms.

In their 2018 paper, S. Wen, W. Zhou, J. Zhang, Y. Xiang, W. Zhou, and W. Jia analyze that social network worms, such as email worms and facebook worms, pose a critical security threat to the Internet. Modeling their propagation dynamics is essential to predict their potential damages and develop countermeasures. Although several analytical models have been proposed for modeling propagation dynamics of social network worms, there are two critical problems unsolved: temporal dynamics and spatial dependence. First, previous models have not taken into account the different time periods of Internet users checking emails or social messages, namely, temporal dynamics. Second, the problem of spatial dependence results from the improper assumption that the states of neighbouring nodes are independent. These two problems seriously affect the accuracy of the previous analytical models. To address these two problems, we propose a novel analytical model. This model implements a spatial- temporal synchronization process, which is able to capture the temporal dynamics. Additionally, we find the essence of spatial dependence is the spreading cycles. By eliminating the effect of these cycles, our model overcomes the computational challenge of spatial dependence and provides a stronger approximation to the propagation dynamics. To evaluate our susceptible-infectious-immunized (SII) model, we conduct both theoretical analysis and extensive simulations. Compared with previous epidemic models and the spatial-temporal model, the experimental results show our SII model achieves a greater accuracy. We also compare our model with the susceptible-infectious-susceptible and susceptible-infectious-recovered models. The results show that our model is more suitable for modeling the propagation of social network worms.

**Existing System:**

Brodka, Mateusz Sobas and Henric Johnson in their paper have proposed two novel methods for detecting cloned profiles. The first method is based on the similarity of attribute values from original and cloned profiles and the second method is based on the network relationships. A person who doubts that his profile has been cloned will be chosen as a victim. 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 is his original profile and which one is a duplicate.

Cresci S, Di Pietro R, Petrocchi M, Spognardi A, Tesconi M, in their paper have reviewed some of the most relevant existing features and rules (proposed by Academia and Media) for fake Twitter accounts detection. They have used these rules and features to train a set of machine learning classifiers. Then they have come up with Class A classifier which can effectively classify original and fake accounts.

**Proposed System:**

In the proposed system we used to detect fake Twitter profiles. Here fake profiles are detected based on rules that effectively distinguish fake profiles from genuine ones. Some of the rules that are used to detect fake profiles are - usually fake profiles do not have profile name or image.

They do not include any description about the account. The geo-enabled field will be false as they do not want to expose their location in tweets.

They usually make large number of tweets or sometimes the profiles would not have made any tweets etc.

The rules are applied on the profile, for each matching rule, a counter is incremented, if the counter value is greater than pre-defined threshold, then the profile is termed as fake.

**Proposed System Advantage:**

- The modules worked fine and was able to detect clones with good accuracy.
- Good Results

**System Architecture**

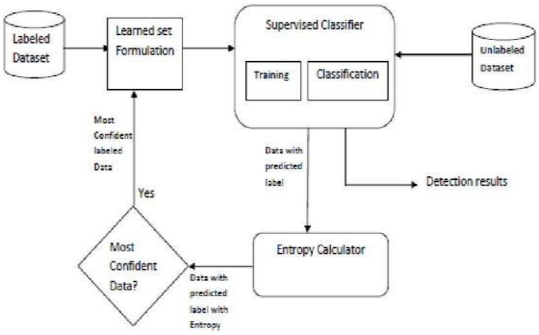


Figure 1: System Architecture

## III. METHODOLOGY

The system is organized into key modules, each designed to handle distinct aspects of the fake user identification on social network.The modules are as follows:

- Data Collection
- Dataset
- Data Preparation
- Model Selection
- Analyze and Prediction
- Accuracy on test set
- Saving the Trained Model

### 3.1.2 Module Descriptions

**1. Data Collection**

This is the first real step towards the real development of a machine learning model, collecting data. This is a critical step that will cascade in how good the model will be, the more and better data that we get, the better our model will perform.

**2. Dataset**

The dataset consists of 1338 individual data.There are 9 columns in the dataset,which are described below:

**ID**: Id number<br>
**UserID**: twitter id<br>
**No Of Abuse Report**: The number of Abuse Report<br>
**No Of Rejected Friend Requests**: The number of Rejected Friend Requests  Followers in the  twitter amount<br>
**No Of Friends**: The number of people friends in the twitter amount<br>
**No Of Followers**: The number of people Followers in the twitter amount<br>
**No Of Likes To Unknown Account**: The number of Likes To Unknown Account<br>
**No Of Comments Per Day**: The number of Comments Per Day<br>
**Fake Or Not Category**: 1 OR 0

**3. Data Preparation**

we will transform the data. By getting rid of missing data and removing some columns. First we will create a list of column names that we want to keep or retain. Next we drop or remove all columns except for the columns that we want to retain. Finally we drop or remove the rows that have missing values from the data set.

**4. Model Selection**

While creating a machine learning model, we need two dataset, one for training and other for testing. But now we have only one. So lets split this in two with a ratio of 80:20. We will also divide the dataframe into feature column and label column. Here we imported train_test_split function of sklearn. Then use it to split the dataset. Also, test_size = 0.2, it makes the split with 80% as train dataset and 20% as test dataset. The random_state parameter seeds random number generator that helps to split the dataset. The function returns four datasets. Labelled them as train_x, train_y, test_x, test_y. If we see shape of this datasets we can see the split of dataset.

**5. Analyze and Prediction**

In the actual dataset, we chose only 7 features :

**UserID**: twitter id<br>
**No Of Abuse Report**: The number of Abuse Report<br>
**No Of Rejected Friend Requests**: The number of Rejected Friend Requests Followers in the twitter amount<br>
**No Of Friends**: The number of people friends in the twitter amount<br>
**No Of Followers**: The number of people Followers in the twitter amount<br>
**No Of Likes To Unknown Account**: The number of Likes To Unknown Account<br>
**No Of Comments Per Day**: The number of Comments Per Day

**6. Accuracy on test set:**

We got a accuracy of 95.1% on test set.

**7. Saving the Trained Model:**

Once you're confident enough to take your trained and tested model into the production-ready environment, the first step is to save it into a .h5 or . pkl file using a library like pickle .

Make sure you have pickle installed in your environment.

Next, let's import the module and dump the model into . pkl file3.2 Technique

We will use Random Forest Classifier, which fits multiple decision tree to the data. Finally I train the model by passing train_x, train_y to the fit method. Once the model is trained, we need to Test the model. For that we will pass test_x to the predict method. Random Forest is one of the most powerful methods that is used in machine learning for classification problems. The random forest comes in the category of the supervised regressor algorithm. This algorithm is carried out in two different stages the first one deals with the creation of the forest of the given dataset, and the other one deals with the prediction from the regressor.

## IV. IMPLEMENTATION

The system is implemented in web environment using struts framework. The apache tomcat is used as the web server and windows xp professional is used as the platform. Interface the user interface is based on Struts provides HTML Tag.

This project is implements like web application using Python and the Server process is maintained using the SOCKET & SERVERSOCKET and the Design part is played by Cascading Style Sheet.

## V. EXPERIMENTAL RESULTS

**Home page**

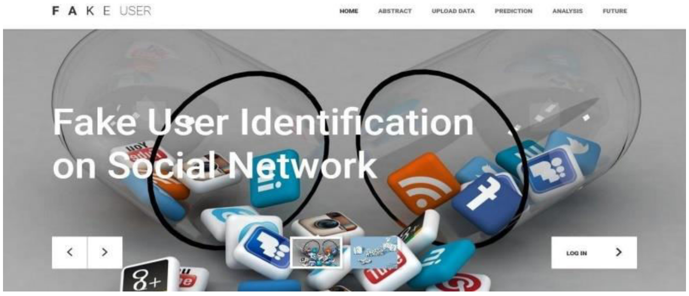


*Figure 2:Home Page*

The homepage Contains login.

**Login Page**

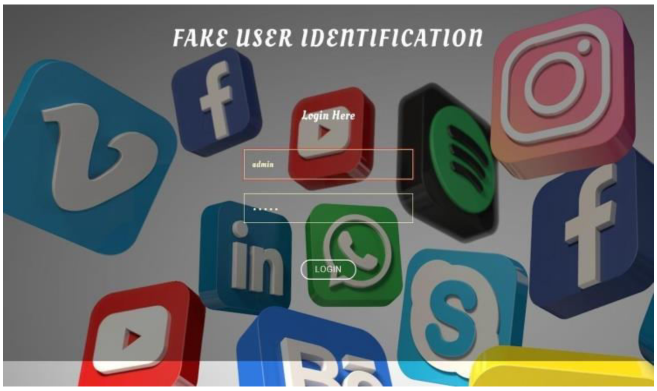
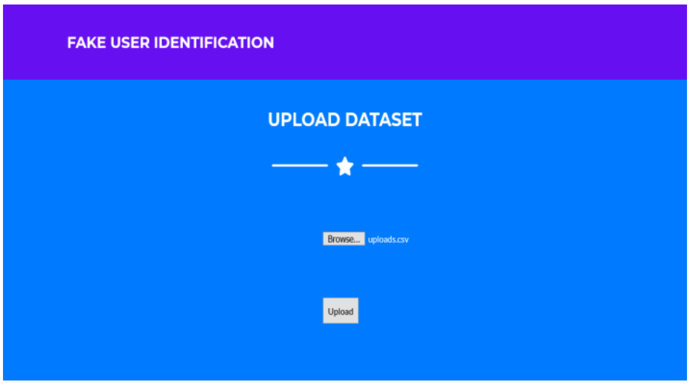


*Figure 3:Login Page*

In the login page user has to enter his/her username and password to login to the page.

### Uploading Datasets

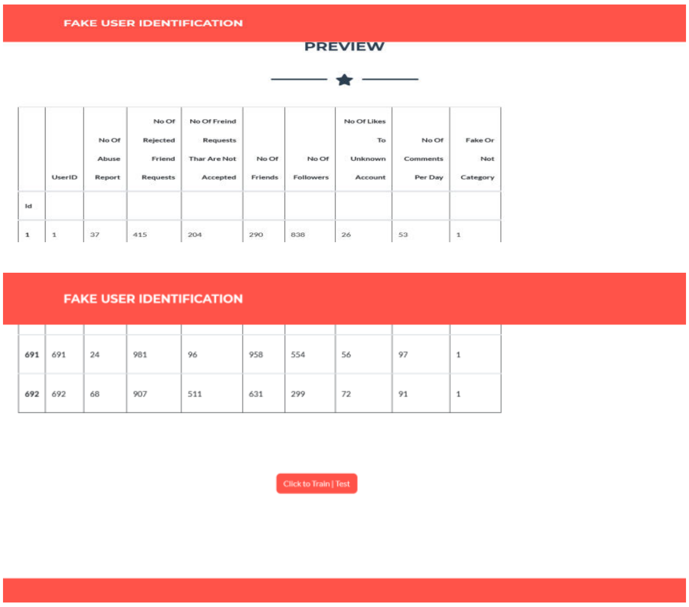


FAKE USER IDENTIFICATION

UPLOAD DATASET

*Figure 4:Uploading Datasets*

### Dataset Preview Page



FAKE USER IDENTIFICATION

PREVIEW

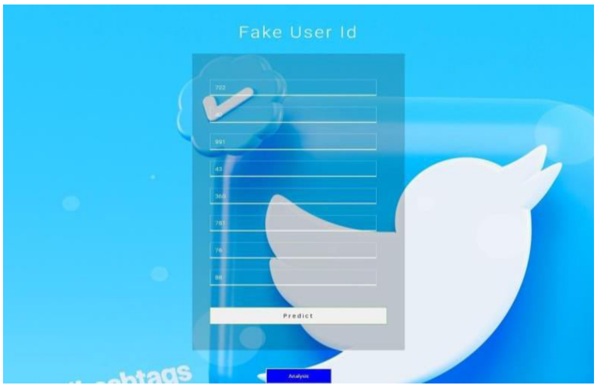
|  | UserID | No Of Abuse Report | No Of Rejected Friend Requests | No Of Freind Requests Thar Are Not Accepted | No Of Friends | No Of Followers | No Of Likes To Unknown Account | No Of Comments Per Day | Fake Or Not Category |
|---|---|---|---|---|---|---|---|---|---|
| Id |  |  |  |  |  |  |  |  |  |
| 1 | 1 | 37 | 415 | 204 | 290 | 838 | 26 | 53 | 1 |

FAKE USER IDENTIFICATION

|  |  |  |  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|---|---|---|
| 691 | 691 | 24 | 981 | 96 | 958 | 554 | 56 | 97 | 1 |
| 692 | 692 | 68 | 907 | 511 | 631 | 299 | 72 | 91 | 1 |

*Figure 5:Dataset Preview Page*

### Data Entry



Fake User Id

702

40

991

43

360

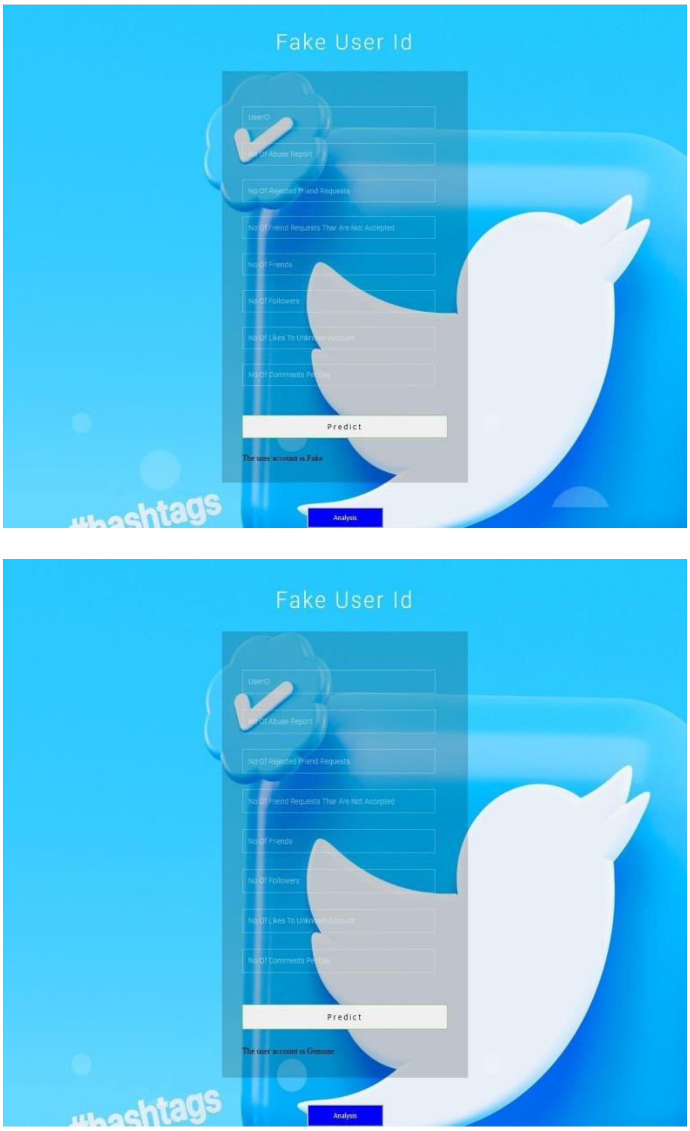
781

76

98

*Figure 6:Data Entry*

**Data Prediction**



The user account is Fake

The user account is Genuine

*Figure 7:Data Prediction*

This prediction page shows whether the account isgenuine or fake.

## VI. CONCLUSION

Fake and clone profiles have become a very serious problem in online social networks. We hear some or the other threats caused by these profiles in everyday life. So a detection method has been proposed which can find both fake and clone Twitter profiles. For fake detection, a set of rules were used which when applied can classify fake and genuine profiles. Clone detection was carried out using Similarity Measures and C4.5 algorithm and a comparison was made to check the performance. Clone detection using Similarity Measures worked better than C4.5 and was able to detect most of the clones which were fed into the system.

## VII. FUTURE ENHANCEMENT

They usually make large number of tweets or sometimes the profiles would not have made any tweets etc. The rules are applied on the profile, for each matching rule, a counter is incremented, if the counter value is greater than pre-defined threshold, then the profile is termed as fake. This module detects clones based on Attribute and Network similarity.

User profile is taken as input. User identifying information are extracted from the profile. Profiles which are having attributes matching to that of user's profile are searched. Similarity index is calculated and if the similarity index is greater than the threshold, then the profile is termed as clone, else norma.In this work we have considered only the profile attributes for fake and clone detection. In future this work can be extended by taking tweets also into consideration by applying some NLP techniques

## REFERENCES

1. Sowmya P and Madhumita Chatterjee ," Detection of Fake and Cloned Profiles in Online Social Networks", Proceedings 2019: Conference on Technologies for Future Cities (CTFC)
2. Georgios Kontaxis, Iasonas Polakis, Sotiris Ioannidis and Evangelos P.Markatos, "Detecting Social Network Profile Cloning", 2013
3. Piotr Bródka, Mateusz Sobas and Henric Johnson, "Profile Cloning Detection in Social Networks", 2014 European Network Intelligence Conference
4. Stefano Cresci, Roberto Di Pietro, Marinella Petrocchi, Angello Spognardi, Maurizio Tesconi, "Fame for sale: Efficient detection of fake Twitter followers", 2015 Elsevier's journal Decision Support Systems, Volume 80
5. 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
6. M.A.Devmane and N.K.Rana, "Detection and Prevention of Profile Cloning in Online Social Networks", 2014 IEEE International Conference on Recent Advances and Innovations in Engineering
7. Kiruthiga. S, Kola Sujatha. P and Kannan. A, "Detecting Cloning Attack in Social Networks Using Classification and Clustering Techniques" 2014 International Conference on Recent Trends in Information Technology

.

IJARETY

International Journal of Advanced Research in Education and Technology

ISSN: 2394-2975 Impact Factor: 7.394

www.ijarety.in editor.ijarety@gmail.com