



ISSN: 2454-9940



**INTERNATIONAL JOURNAL OF APPLIED
SCIENCE ENGINEERING AND MANAGEMENT**

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

Detection of Fake and Clone accounts in Twitter using Classification and Distance Measure Algorithms

¹B.N.S. GUPTA, ²GOPISETTI LAKSHMI LAVANYA

¹(Assistant Professor), MCA, S.V.K.P & Dr K.S. Raju Arts & Science College

²MCA, scholar, S.V.K.P & Dr K.S. Raju Arts & Science College

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.

1.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 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 paper organized as below. Section II describes the literature survey. Section III explains the proposed methodology. Section IV discusses the results. At last, Section V concludes the paper with the conclusion.

2.LITERATURE SURVEY

Currently, existing works make it difficult to investigate Twitter spam, especially element selection and location calculations.

1) When determining emphasis, genealogy often chooses an indistinguishable type of quality.[1]Content-based and customer profile-based attributes for detection. Many types of attributes in the informal community of rare customers are unique compared to the attributes of common customers, so it is not enough to accurately convey the state of the information.

2) In [8] calculation decisions, analysts primarily use AI calculations to manage the location of spam in interpersonal organizations. Considering the possibility of characterization, scientists have provided mathematical structural features to distinguish spam clients.

3) The actual dataset of the informal organization shows the long tail effect. H. This is a heterogeneous dataset that contains a wide variety of non-spam that far outweighs spam.[10] Performance suffers when these managed AI calculations are recognized by non-uniform datasets. Similarly, you need to take advantage of multi-tiered attributes and perform calculations prepared to show end-to-end reachability even if the dataset is nonuniform.

3. EXISTING SYSTEM

Georgios Kontaxis, Iasonas Polakis, Sotiris Ioannidis and Evangelos P Markatos [2] have proposed a prototype to check whether the users have become victim to cloning attack or not. Information is extracted from user profile and a search is made in OSN to find profiles which

match to that of user profile and a similarity score is calculated based on commonality of attribute values. If the similarity score is above the threshold value then the particular profile is termed as clone.

Brodka, Mateusz Sobas and Henric Johnson in their paper [3] 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 $S(Pc, Pv) > \text{Threshold}$, 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 [4], 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.

Ahmed El Azab, Amira M Idrees, Mahmoud A Mahmoud, Hesham Hefny [5], have proposed a classification method for detecting fake accounts on Twitter. They have collected some effective features for the detection process from different research and have filtered and weighted them in first stage. Various experiments are conducted to get minimum set of attributes which gives accurate results. From 22 attributes, only seven attributes were selected which can effectively detect fake accounts and have applied these factors on classification techniques. A comparison of the classification techniques based on results are made and the one which provides most accurate result is selected.

➤ **Disadvantages**

- In the existing work, the system doesn't calculate fake accounts due to lack of Attribute similarity finding.
- This system less effective due to absence of Attribute similarity which is not calculated based on the similarity of attribute values between the profiles.

3.1 PROPOSEDSYSM

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. The architecture of proposed system is as shown in the proposed system.

The proposed architecture consists of modules for Fake Profile detection and Clone Profile detection.

A. Fake Profile Detection

This module is 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.

➤ B. Clone Profile Detection using Similarity Measures

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 normal[1].

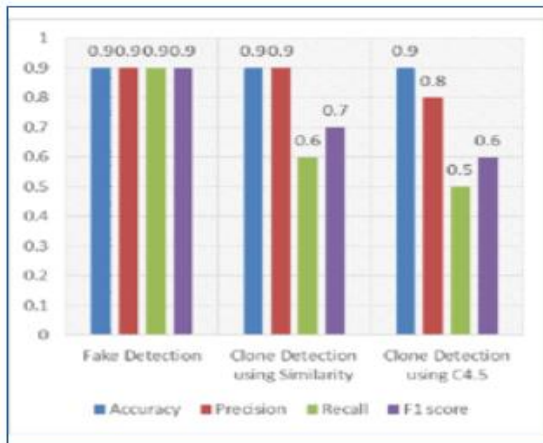
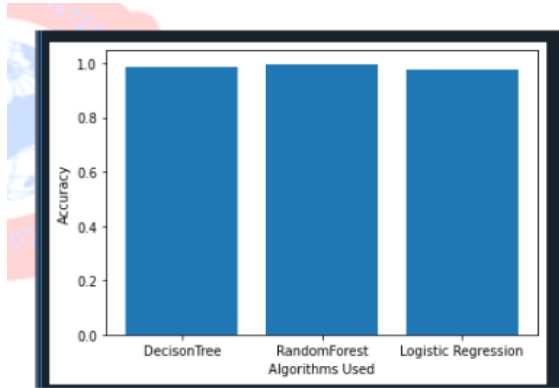
➤ i) Attribute Similarity

➤ Attribute similarity is calculated based on the similarity of attribute values between the profiles. The attributes that are considered for similarity measurement are Name, ScreenName, Language, Location and Time_zone. Two similarity measures are used to measure the similarity between the attributes – Cosine similarity and Levenshtein distance. Cosine similarity is used to find similarity between words and Levenshtein distance is used to find similarity between two sequences.

➤ Advantages

- Accuracy which gives the ratio of number of correct results to the total number of inputs.
- Precision which gives the proportion of positive detection that was actually correct.
- Recall which gives the proportion of actual positives that was detected correctly.

4. OUTPUTSCREENS



5. 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.

6. REFERENCE

- [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 Brodka, 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: Efficientdetection 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