Section: Research Paper

Study and Analysis of Types of social media Fake profiles and Machine Learning Algorithms for Detection and classification of Fake Profiles

Maulik Shah[1] and Hiren Joshi[2]

[1] *Department of computer science, Gujarat university, Ahmedabad, India*

[2] *Department of computer science, Gujarat university, Ahmedabad, India*

Abstract. The use of OSNs by persons in today's society for the purpose of participating in their typical social activities is becoming more common. As a consequence of this, a considerable quantity of data pertaining to the users' social lives, personal lives, and professional lives is being stored on these OSNs. In spite of the fact that using these sites has led to an overall improvement in the quality of people's social lives, one of the disadvantages that is linked with using them is the proliferation of fake accounts. Utilization of these products is connected to a variety of other issues. Academics and social analysts are drawn to the user data that is made public on open social networks (OSNs), but hackers are also interested in this data since it may be exploited in many ways. These cybercriminals take advantage of the openness and susceptibility of an OSN by generating fake identities, which they then use to engage in activity that is unlawful, dishonest, and damaging. Theft of identity, slandering, trolling, bullying, and spamming are some examples of the behaviors that fall under this category. To transmit spam, conduct fraud, or otherwise abuse the system in any other manner, criminal users of online social networks like the approach of establishing false accounts, which is the favored method. Users who have created false identities in order to participate in criminal operations have established themselves on the most major social networking sites. In this paper we have described the different types of social media fake profiles and hybrid algorithm for detection and classification of fake Instagram profile.

Keywords: Online Social Network, Neural Network, Decision tree, K-Nearest Neighbor (KNN)

1 Introduction

The use of OSNs by persons in today's society for the purpose of participating in their typical social activities is becoming more common. As a consequence of this, a considerable quantity of data pertaining to the users' social lives, personal lives, and professional lives is being stored on these OSNs. In spite of the fact that using these sites has led to an overall improvement in the quality of people's social lives, one of the disadvantages that is linked with using them is the proliferation of fake accounts. Utilization of these products is connected to a variety of other issues. Academics and social analysts are drawn to the user data that is made public on open social networks (OSNs), but hackers are also interested in this data since it may be exploited in many ways. These cybercriminals take advantage of the openness and susceptibility of an OSN by generating fake identities, which they then use to engage in activity that is unlawful, dishonest, and damaging. Theft of identity, slandering, trolling, bullying, and spamming are some examples of the behaviors that fall under this category. To transmit spam, conduct fraud, or otherwise abuse the system in any other manner,

criminal users of online social networks like the approach of establishing false accounts, which is the favored method. Users who have created false identities in order to participate in criminal operations have established themselves on the most major social networking sites. According to a report1, the most prominent social networking site, Facebook, has recognized and erased more than 580 million Facebook identities from the channel in the year 2018, while more than 87 million phoney accounts are still active on the platform [1]. According to another finding in the survey, there are over 87 million fraudulent accounts now active on the network. Twitter has reportedly deleted more than 70 million fake accounts in 2018, according to a separate report2. 3 In addition, there are more than 45 million fake accounts, which represent more than 15 percent of the total number of monthly active users on the network. One of the most major challenges in this day and age, while using the internet, is the abundance of fake accounts that can be found online. Because these criminals use open-source networks as a tool to perform daily significant crimes, cyber intelligence is having a lot of difficulties finding ways to lessen the effect that these profiles have. According to allegations by NDTV4, a gang of Iranian hackers constructed around 14 bogus profiles across a range of online social networks (OSNs), one of which being Facebook, with the purpose of monitoring various components of the political and military establishment in the United States. They were effective in misleading those folks because they established friend relationships with around 2,000 of the other users on the network. In the beginning, the hackers did not provide victims any information that may be considered dangerous in order to build the victims' trust. After that, the cybercriminals used bogus identities to send links to the computers of the victims, which would install dangerous malware on those machines. Another study [2] came to the conclusion that Facebook had uncovered and deactivated 32 fake accounts that were a part of a strategy to give the appearance of having undue political influence. It has been discovered that these accounts were created between the months of March 2017 and May 2018, according to the information that has been uncovered. Several instances of people's identities being stolen and exploited to create bogus accounts online were described in depth by NBC News5 in a series of reports that were broadcast. As an example, only recently, one of the men who serves on the Atlanta City Council, Alex Wan, discovered that his photo was being used by a number of fake identities in an effort to attract women.

Scientists have devised a wide number of tactics and put them into practice in order to recognize, counteract, and lessen the influence that these fake profiles have on OSNs. However, attackers are skilled at finding new methods to bypass these defenses so that they may continue to fool the network. This allows them to continue their malicious activity. To put an end to the problem of phoney profiles being used on OSNs, there has to be an efficient way for identifying phoney profiles on OSNs. This is essential in order to put an end to the problem.

2 Literature Review

An online social network, or OSN, is defined as a collection of nodes that are connected to one another via a network of connections. Nodes may represent persons, actors, organizations, nations, and governments, among other entities (interactions, hyperlinks, etc.). OSNs, which stands for "online social networks," are web-based apps that primarily aim to encourage user participation and collaboration, in addition to the exchange of information. This is the primary focus of the programmes that fall under the OSN umbrella. As a direct consequence of online social networks, people's

modes of thought, methods of expressing themselves, and interactions with the wider world have all been transformed. People currently carry out their professional and social activities by using a range of social networking sites, such as Facebook, Twitter, Flicker, LinkedIn, Scientific databases, and others. Facebook, Twitter, and Research Gate are just a few examples of the websites that fall within this category. Researchers and professionals in a variety of other fields, such as marketing, sociology, politics, and others, place a significant emphasis on online social networks (OSNs) because they have a structure that is comparable to that of real-life communities and because they store a massive amount of user content. The reason for this is that OSNs have a structure that is similar to that of real-life communities. OSNs are examined by marketing firms in order to build viral marketing methods and to attain their intended clientele; sociologists use them in order to get a better understanding of human behavior; and politicians use them in order to boost their political campaigns. [3][4][5].

Online social networks, which provide a potent communication channel, have emerged as popular devices for improving information sharing and boosting social engagement. This is because online social networks provide a strong communication channel. Additionally playing a significant part in world affairs are the many online social networks. In addition to keeping in touch with members of their personal networks, such as friends and family, users of open social networks (OSNs) have the ability to interact with members of professional groups. People join up for OSNs and make connections with other members of the network with the goal of continuing to communicate with those members and having conversations on things that they have in common with those people. One of the most important and basic features that are present across all OSNs is the ability for users to create their own profiles, which is offered by the vast majority of online social networks (OSNs). Users have the ability to post content inside their profiles, which may include text, photographs, videos, and other forms of media. This content, which may be seen by all members of the network or by a specific subset of those users, is referred to as "public." OSNs simplify the process of establishing connections between users and other individuals who are already participating in the network. It is thought that two members who have a link are familiar with one another and get along swimmingly when there is a connection between them (or that they are neighbors). Users that live in close proximity to one another often have same interests.

There is a vast selection of OSNs from which to choose, and which one to use is entirely determined by the criteria. For instance, there are social networks that aid people in building online social ties with members of their families and friends' social networks. These social networks are referred to as "social relationship builders." Some examples of social networks that may be found online are Twitter, Facebook, and Myspace (OSNs). Facebook is by far the most widely used social networking site that can be found on the internet. Users are provided with a platform via which they may interact and exchange information with the individuals they already know. Twitter gives its users the ability to broadcast their thoughts, ideas, and suggestions while also giving them the chance to receive updates from other users who are related to them. YouTube and Flickr are two examples of websites that are examples of social networking websites that were created with the express purpose of making it simple and straightforward to upload and share movies and photographs. YouTube and Flickr were created with the express purpose of making it possible to share movies and photographs. In addition, there are online networking programmes that were developed with the express purpose of supporting the users' personal and professional growth. One example of such a programme is LinkedIn, which was developed with this goal in mind. Because it is an online network of people, each participant has the opportuni-

ty to communicate with others who are engaged in the same area of work, which is a feature that may prove to be very beneficial. One of the many types of online social networks is the discussion forum, which is also one of the forms of online social networks. Users are able to share their knowledge with one another via the usage of discussion forums, such as the one that can be found on Quora6. These forums allow users to ask and answer questions relating to a certain topic with other users. Online social networks (OSNs) have been organized by our team into the following five basic categories: Pure OSNs, Media sharing OSNs, Professional OSNs, Discussion forums and Blogs, and Dating OSNs. These classifications were established on the basis of the functions and conveniences offered to users by the various OSNs. An explanation of each of the aforementioned categories is provided in the following abbreviated form: Websites that individuals use as an essential component of their day-to-day social activities are referred to as pure online social networks, or POSNs for short. These websites also go by the acronym POSN. These activities include establishing new acquaintances, exchanging information, discussing about forthcoming events, joining communities, reading the news, and a great deal of other things as well. POSNs also provide its users tools that facilitate communication, such as audio and video chatting, real-time messaging, and the ability to update a post with the user's current location, amongst other features. When it comes to forming social ties with people over the internet, the POSNs are by far the most popular option. Among the most well-known public online social networks (POSNs) are social networking sites such as Facebook, Twitter, and Sina Weibo.The term "media sharing OSNs" refers to many types of online social networking platforms that, rather than concentrating only on the dissemination of written material, place a primary emphasis on the exchange of purely visual content, such as photographs and videos. There are already a variety of websites available to consumers that enable them to share various types of material. The users of these websites are able to communicate with one another by the simple act of publishing photographs or videos of their activity. In terms of social networks, some examples of those that share media are Snap chat, YouTube, and Instagram. Since the beginning of this decade, a number of social networks have been developed that enable users to exchange images and videos.

A sizable number of people are actively engaged on these websites, and this number is expected to continue to grow. As the number of people using smartphones continues to increase at an alarming pace, more and more members of the general public are being attracted into these sites. These days, users may simply take images of a high quality with their mobile devices, and with the help of a range of programmers, they are also able to edit and publish these photographs on websites that allow media sharing. There are more, more well-known platforms for media sharing that include a range of functionalities, such as, for example.

Researchers in the scientific community devised a variety of strategies and put those strategies into action in order to identify and combat false accounts on online social networks (OSNs), as well as to decrease the harmful effect that these accounts have. However, attackers are competent at discovering new ways to circumvent these safeguards so that they may continue to mislead the network. This allows them to continue their malicious activity. Because of this, they are able to go on with their nefarious conduct. In order to put an end to the issue of phoney profiles being used on OSNs, there has to be a method that is both effective and efficient for recognizing phoney profiles on OSNs. This is very necessary in order to solve the issue once and for all.

3 Types of Fake Profiles in Online Social Networks

On the basis of the characteristics of phoney profiles, we have divided them into the five groups listed below: hacked profiles, clone accounts, sybil accounts, sockpuppet profiles, and fake bots profiles. This breakdown is seen in figure 1. The sub-sections that are to follow each provide a separate study of one of the categories that were covered in the previous section. Opponents pursue their evil goals across a range of online social networking sites using a variety of strategies, which may be classified into these categories and thought of as that many different ways.

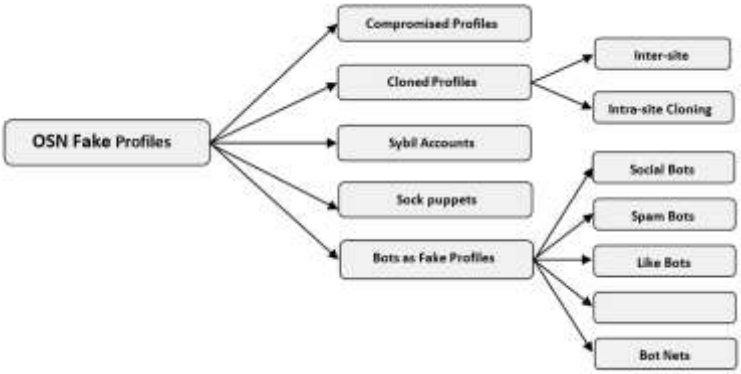

Figure 1:-Difference types of fake profiles in online social

3.1 Compromised Profiles

Compromised accounts are real accounts, but the owners either do not have complete control over them or have lost access to a phisher or another form of malware agent [8]. Compromised accounts may be used to commit identity theft. According to the terms and conditions of Facebook, a genuine account is considered to be compromised if it can be accessed by a person who is not the authorized owner of the account. In other words, if the account can be used by someone other than the owner, the account is considered to be compromised. This is true for all of your accounts. According to the authors who contributed to hijacked accounts, the kind of accounts that are the most difficult to uncover are the ones that were compromised by hackers. This is because the real owner of the account has likely already built up some level of credibility on the internet [9]. According to the findings of a different piece of recent research, out of the total number of detected malicious accounts that were used to disseminate spam, more than 97 percent of the profiles were found to be hacked rather than fictitious [10].

3.2 Cloned Profiles

A technique known as "cloning a profile" is when an adversary establishes a second profile by replicating the data of an existing genuine profile, such as its username, age, nationality, profile image, and so on. This allows the adversary to impersonate the original user of the original profile. To put it another way, we can say that the process of profile cloning is the act of stealing the information of the victim in order to create another profile as a way to spread spam, collect private information from the victim and the victim's friends, or carry out other scams such as stalking, defaming,

and other activities that are very similar to these. In other words, we can say that the process of profile cloning is the act of stealing the information of the victim.

3.3 Sockpuppets Profiles

An online persona known as a "sockpuppet" is an identity that has been fabricated with the intention of fooling other users or of promoting another person or entity on online platforms such as discussion boards, blogs, and social networking sites, amongst other places. This can be done in order to promote another person or entity. Sockpuppets are online accounts that are created with the intention of deceiving internet users in a variety of different ways, such as by leading users to believe that a particular product is a good investment, that there is a low risk associated with an investment plan that has a high return [15], etc. In other words, sockpuppets are online accounts that are created with the intention of deceiving internet users in a variety of different ways. The act of prohibited users on an OSN site creating new accounts in order to continue using the site is referred to as the practise of "sock puppetry," which is a word that describes the term. [16]

3.4 Sybil Accounts

In the case of sybil accounts, dishonest individuals create a large number of accounts in order to launch an attack on a dependable network. These individuals then take control of each individual account in turn. A node in an online social network that fraudulently claims to perform numerous roles and presents a threat to the network's security is referred to as a Sybil attack. This assault term comes from the fact that this kind of circumstance has been given a name. By independently constructing and maintaining a large number of pseudonymous identities over a prolonged period of time, an attacker may undermine the public image of a network. The attacker acquires a disproportionately big level of influence as a consequence of the use of these identities to propagate malware and spam on social networks, which in turn results in the attacker gaining a disproportionately high amount of influence. A Sybil attack is the name given to this particular kind of assault.

3.5 Bots as Fake Profiles

A bot is a piece of software that, after it has been installed on a computer, will perform numerous scripts in order to simulate human behavior on the internet. Bots are often used in online gaming communities. According to the authors of "23," a "bot" is a computer programme that generates certain data in order to communicate with people, specifically internet users (also known as netizens), in order to influence the behavior of such individuals. Bots are used to communicate with people in order to influence their behavior.

Table 1:-Fake Bots and their characteristics

Fake Bots and their characteristics					
	Social Bots	Spam Bots	Like Bots	Influential Bots	Bots net

Purpose	To create social, personal and professional relations	To spread malicious content	To increase the ranking/ratings To gain false followers	To alter the behaviour of people To perform viral marketing	To perform various unlawful operations by a bot network
Used by	Politicians, Researchers Academicians	Cyber criminals, Marketing and advertising agencies.	Marketing and advertising agencies	Politician Marketing and advertising agencies, Celebrities	Researchers, Hackers, Marketing and advertising agencies, etc.
Networks	OSNs, Discussion forums.	OSNs.	e-commerce sites, OSNs.	OSNs, Discussion-forums, e-Commerce.	Discussion forums, OSNs.

4 Research Gap Analysis

On every social networking site, there are people using fake accounts for a variety of reasons. The many characteristics of false profiles and the strategies they use to accomplish their objectives continue to evolve with the passage of time. On the one hand, new detection methods and systems are being devised, while on the other side, adversaries adopt sophisticated tactics in order to dodge from these detection systems. In the sub-sections that follow, we will have a short discussion about some of the most significant difficulties linked with the identification of phoney profiles in OSNs. Here in our research work we are mainly focus on Instagram and Twitter. As majority of the fake profile on this platforms are high as compared to others .The main research gap identified are

1. Many state-of-the-art approaches, including neural networks, KNN, and decision trees, have been shown to have poor accuracy and a high error rate.
2. The use of high-dimensional data by machine learning algorithms leads to overfitting, which isn't a good thing for either party.
3. The presence of hybrid characteristics in the same document results in an incorrect evaluation.

Other details of research gaps are given below:

4.1 Lack of access to ground-level information Data and effective instruments for the collection of data: A significant number of data is the preliminary need for carrying out any form of analysis. Researchers may cut down on their workload and save a significant amount of time by making use of ground-truth databases. Additionally, it has a lower amount of noise, outliers, and missing data, among other things, and as a result, it does not need stringent pre-processing. Since we are mainly concerned with false profiles, we have discovered that the datasets that are now available do not include nearly all of the characteristics that are necessary for the development of detection algorithms for phoney profiles. In addition, there is not yet a benchmark dataset that has been made available for phoney accounts on Facebook.

4.2 Putting together fake profiles in preparation for model training: We require a training dataset that is enriched with both actual and false profiles so that we may build a machine learning-based fake profile identification system. Because there are a large number of genuine accounts now accessible on the social network, it is not diffi-

cult to find them. One method for gathering actual profiles is to leverage the accounts of reputable users as well as their personal networks of acquaintances. On the social network, trusted people might be friends, verified accounts, or profiles of persons who are already known to the user directly. When compared to actual profiles, however, the number of false accounts is far lower. Because of this, it is very difficult for researchers to find phoney accounts on a massive network that has billions of members.

4.3 The most effective feature set for identifying fake profiles: In order to identify phoney profiles on social networking websites, the first thing that has to be done is to conduct an analysis of the features that set these profiles apart from the genuine ones. In order to construct a false profile detector that works well, one must first create a feature set that is relevant, accurate, and efficient. You may either personally examine the characteristics on the social network sites or investigate them using a literature study. Both options are available. However, it is also feasible that some of the characteristics described in the current literature will not prove to be productive in the present context. This is due to the fact that adversaries are constantly altering their behavior in order to mislead and get around detection systems. Several academics have, on occasion, taken the effort to identify various aspects of online profiles in order to better train their models designed to detect phoney accounts. Researchers have classified features into a number of different categories, each of which is determined by the nature of the feature itself.

5 Conclusion

In this research paper we have presented the different approach of the fake profile creation and usage. This paper also covers the different types of the social media fake profile with its pros and cons. While writing and during this research we have identified many research gap which we have presented in the research paper. This paper also contains the research gap which we have analyzed for the fake profile identification of Instagram accounts. This research paper contributes in the field of fake profile identification of Instagram as well the profile explained in the types of profile section helps the user to identify what type of the profile it is what are its consequences.

References

1. Andrew Hutchinson, “Facebook Outlines the Number of Fake Accounts on Their Platform in New Report,” 2018. [Online]. Available: https://www.socialmediatoday.com/news/facebook-outlines-the-number-of-fakeaccounts-on-their-platform-in-new-repo/523614/. [Accessed: 14-Jan-2019].
2. K. R. Nicholas Fandos, “Facebook Identifies an Active Political Influence Campaign Using Fake Accounts - The New York Times,” The New York Times, 2018. [Online]. Available: https://www.nytimes.com/2018/07/31/us/politics/facebook-political-campaignmidterms.html. [Accessed: 18-Jan-2019].
3. M. Vergeer, L. Hermans, and S. Sams, “Online social networks and microblogging in political campaigning,” Party Polit., vol. 19, no. 3, pp. 477–501, May 2013.
4. S. Staab et al., “Social Networks Applied,” IEEE Intell. Syst., vol. 20, no. 1, pp. 80–93, Jan. 2005.

5. N. A. Christakis and J. H. Fowler, "Social contagion theory: examining dynamic social networks and human behavior," Stat. Med., vol. 32, no. 4, pp. 556–577, Feb. 2013.
6. T. Aichner, "Measuring the degree of corporate social media use," no. March, 2018.
7. K. R. Nicholas Fandos, "Facebook Identifies an Active Political Influence Campaign Using Fake Accounts - The New York Times," The New York Times, 2018. [Online]. Available: https://www.nytimes.com/2018/07/31/us/politics/facebook-political-campaignmidterms.html. [Accessed: 18-Jan-2019].
8. D. Savage, X. Zhang, X. Yu, P. Chou, and Q. Wang, "Anomaly detection in online social networks," Soc. Networks, vol. 39, no. 1, pp. 62–70, 2014.
9. M. Egele, C. Kruegel, and G. Vigna, "COMPA : Detecting Compromised Accounts on Social Networks," 2016.
10. H. Gao, J. Hu, C. Wilson, Z. Li, Y. Chen, and B. Y. Zhao, "Detecting and characterizing social spam campaigns," Proc. 10th Annu. Conf. Internet Meas. - IMC '10, p. 35, 2010.
11. M. Y. Kharaji, F. S. Rizi, and M. R. Khayyambashi, "A New Approach for Finding Cloned Profiles in Online Social Networks," vol. 6, no. April, pp. 25–37, 2014.
12. T. Stein, E. Chen, and K. Mangla, "Facebook immune system," Proc. 4th Work. Soc. Netw. Syst. - SNS '11, vol. m, pp. 1–8, 2011.
13. L. Bilge, T. Strufe, D. Balzarotti, and E. Kirda, "All your contacts are belong to us," Proc. 18th Int. Conf. World wide web - WWW '09, p. 551, 2009.
14. B. Bhumiratana, "A model for automating persistent identity clone in online social network," Proc. 10th IEEE Int. Conf. Trust. Secur. Priv. Comput. Commun. Trust. 2011, 8th IEEE Int. Conf. Embed. Softw. Syst. ICESS 2011, 6th Int. Conf. FCST 2011, pp. 681–686, 2011.
15. X. Zheng, Y. M. Lai, K. P. Chow, L. C. K. Hui, and S. M. Yiu, "Sockpuppet detection in online discussion forums," Proc. - 7th Int. Conf. Intell. Inf. Hiding Multimed. Signal Process. IIHMSP 2011, pp. 374–377, 2011.
16. M. Tsikerdekis and S. Zeadally, "Multiple Account Identity Deception Detection in Social Media Using Nonverbal Behavior," Inf. Forensics Secur. IEEE Trans., vol. 9, no. 8, pp. 1311–1321, 2014.
17. S. Kumar, J. Cheng, J. Leskovec, and V. S. Subrahmanian, "An Army of Me: Sockpuppets in Online Discussion Communities," 26th Int. Conf. World Wide Web, pp. 857–866.