IJITCE

ISSN 2347- 3657

International Journal of Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 12, Issue 3, 2024

FAKE PROFILE IDENTIFICATION ON SOCIAL NETWORK USING MACHINE LEARNING AND NLP

[1]K SUPARNA, [2]P. MADHAVI

[1](Assistant Professor), MCA, **DANTULURI NARAYANA RAJU COLLEGE(A) PG COURSES, BHIMAVARAM ANDHRA PRADESH**

[2]MCA, scholar, **DANTULURI NARAYANA RAJU COLLEGE(A) PG COURSES, BHIMAVARAM ANDHRA PRADESH**

ABSTRACT

At present social network sites are part of the life for most of the people. Every day several people are creating their profiles on the social network platforms and they are interacting with others independent of the user's location and time. The social network sites not only providing advantages to the users and also provide security issues to the users as well their information. To analyze, who are encouraging threats in social network we need to classify the social networks profiles of the users. From the classification, we can get the genuine profiles and fake profiles on the social networks. Traditionally, we have different classification methods for detecting the fake profiles on the social networks. But, we need to improve the accuracy rate of the fakeprofile detection in the social networks. In this paper we are proposing Machine learning and Natural language Processing (NLP) techniques to improve the accuracy rate of the fake profiles detection. We can use the Support Vector

ISSN 2347–3657

Volume 12, Issue 3, 2024

Machine (SVM) and Naïve Bayes algorithm.

1.INTRODUCTION

Presently, social networking has grown into a popular online pastime, drawing in hundreds of thousands of users who spend billions of minutes in such sites. There is a wide range of online social network (OSN) services available today, from those that focus on social interactions like Facebook or MySpace to those that prioritize knowledge dissemination like Twitter or Google Buzz, and even those that bring social interaction features to existing systems like Flicker. On the other side, a major obstacle and perceived objective is expanding security measures and safeguarding OSN privacy. Every user of a social network (SN) divulges a unique amount of personal information. Our personal information is vulnerable to several forms of attack, the most serious of which might be identity theft, since it is either completely or partially

exposed to the public. Theft of identity occurs when an unauthorized person exploits another person's knowledge or expert witness for their own gain. Because it impacted millions of individuals all over the globe, online identity theft was a major issue in the past. Individuals who fall prey to identity theft may face a variety of consequences, such as financial loss, loss of time and money, incarceration, harm to their reputation, and loss of connections with friends and family. Many social networks now do not check the debts of regular members and have weak privacy and security regulations. The reality is that the majority of SN apps have very low levels of privacy by default, making them an ideal environment for fraud and abuse. Online social media platforms have made it easier for both experienced and inexperienced criminals to commit identity theft and impersonation assaults. Worse still, while creating an account on a social networking website, users are expected to provide accurate consent. It would be disastrous enough to lose everything if consumers' online activity could be easily monitored; the prospect of such bills being compromised would be devastating. Like offline profiles, online network information may be either static or dynamic. There is static knowledge, which refers to the information that a person may provide when creating a profile, and dynamic knowledge, which refers to the details that

ISSN 2347–3657

Volume 12, Issue 3, 2024

the system inside the network communicates. Differences between static and dynamic knowledge include a person's hobbies and demographics as well as their runtime behaviors and network location. Current research heavily relies on both static and dynamic data. But this doesn't matter on most social networks since users only view a subset of static profiles and dynamic profiles aren't always evident to the person networking. For the purpose of detecting false identities and harmful information in online social networks, many methods have been suggested by diverse researchers. Various procedures have their advantages and disadvantages. Security concerns, abuse, harassment, and trolls are just a few of the numerous issues plaguing social media. Useful for a large number of fake accounts on social media. Blank or generalized profiles are known as false profiles. These profiles include individuals who have provided fraudulent credentials. False Facebook accounts are more likely to engage in harmful and unwanted actions, which may disrupt social community consumers' experience. People make up profiles to promote and advocate for characters or groups of people, engage in social engineering, or to slander another person via online impersonation. Facebook has its own security mechanism in place to protect user credentials from various forms of spam and phishing. An analogous concept is the Facebook Immune System (FIS).More often than not, the FIS has been unable to detect Facebook user- generated bogus accounts.

2.LITERATURE SURVEY

1)" Understanding User Profiles on Social Media for Fake News Detection"

AUTHOR: Kai Shu, Suhang Wang, Huan Liu – 2018

The number of people who get their news via social media is growing rapidly in recent years. Users gain from social media because of its inherent characteristics of rapid distribution, low cost, and simple access. The news is of inferior quality compared to established news , leading to a significant volume of false news. As the negative impacts on people and society grow, the need of detecting false news grows. The current state of false news identification based just on content is sometimes inadequate, hence it is recommended to use

ISSN 2347–3657

Volume 12, Issue 3, 2024

user social interactions as supplementary data to enhance this area. Because of this, it is critical to comprehend the relationship between social media accounts and disinformation in great detail. This work presents the development of real-life datasets that assess the degree to which consumers trust fake news. It also comprises the selection of representative categories of users, namely "experienced" users who can identify falsehoods in fake news items and "naïve" users who are prone to believing such things. We find that these user groups may be distinguished from one another by comparing their explicit and implicit profile traits, which can identify false news. Future research on automated false news identification may build on the results of this work.

2) "Identifying Fake Profiles in LinkedIn"

AUTHOR: Featuring Shalinda Adikari and Kaushik Dutta -

There is growing importance in getting one's profile visible on professional networks like LinkedIn, since more and more companies depend on these sites to create business relationships. This score is directly proportional to the desire to engage in immoral behavior on the network. 3 Building a relationship with someone whose profile is full of false material may be a huge waste of time and energy, and it can also damage the network's credibility overall. True profiles might be hard to spot when they are fraudulent. Though solutions have been suggested for other social media platforms, LinkedIn accounts do not often have their associated data made public. In this study, we assess what information is essential for detecting false profiles on LinkedIn and suggest a data mining strategy for this purpose. Results acquired using bigger data sets and more extensive profile information are similar to our approach's ability to detect false profiles with 87% accuracy and 89% True Negative Rate, even when working with limited profile data. Plus, our solution improves accuracy by around 14% compared to alternatives employing comparable volumes and kinds of data.

3) "A Feature Based Approach to Detect Fake Profiles in Twitte"

AUTHOR:2019—Jyoti Kaubiyal, Ankit Kumar Jain

The popularity of social media sites, especially Facebook and Twitter, has skyrocketed in the

ISSN 2347–3657

Volume 12, Issue 3, 2024

last decade, drawing in millions of users. Many bad actors, including spammers, have taken an interest in them because they have become a popular form of communication. Fake accounts have become more of an issue due to the increasing amount of social media users. False and fraudulent identities are heavily engaged in harmful behaviors including spamming, spreading abuse and disinformation, and artificially increasing an application's user count to advocate for or influence public opinion. Therefore, it is crucial to detect these false identities in order to safeguard legitimate users from harmful intentions. We plan to combat this by using a feature-based strategy to detect these phony accounts on various social networking sites. Our effective identification of bogus accounts is based on twenty-four indicators. For the purpose of confirming the categorization outcomes, three classification

4) "Method for detecting spammers and fake profiles in social network"

AUTHOR: In 2019, Yuval Elovici, Michael FIRE, and Gilad Katz

An approach to safeguarding user privacy in an online social network that uses the database of current members to choose negative examples of fraudulent accounts and positive examples of honest profiles. Next, we extract a set of traits that we know will be useful for identifying phony and real profiles by grouping the friends and followers of our instances into communities and looking at the connections inside and between them. By comparing the properties of existing fake profiles, classifiers trained using supervised learning may identify them.

5) "Social Networks Fake Profiles Detection Using Machine Learning Algorithms"

AUTHOR: Elyusufi twins Yasyn and Zakaria - 2020

A variety of harmful actions, including advanced persistent threats, utilize fake accounts. Finding phoney accounts on social media is the main topic of this article. There are a few different ways to go about detecting false profiles on social media. One way is to look at the data associated with profiles, and another is to study individual accounts. Among all types of cybercrime, the most damaging is the establishment of false profiles on social networks. Notifying the user of the establishment of a false profile is not enough time to discover this crime. There have been several proposals in the literature for algorithms and approaches

ISSN 2347–3657

Volume 12, Issue 3, 2024

that may identify false profiles. 5 By discussing the aforementioned methods for identifying false social media accounts, this article clarifies the function of false identities in APTs. To determine whether a profile is phony or real, we will evaluate the effectiveness of three supervised machine learning algorithms: Random Forest (RF), Decision Tree (DT-J48), and Naïve Bayes (NB).

3. EXISTING SYSTEM

Millions of people use social networking sites like Twitter and Facebook, and their involvement with these sites has a positive impact on their lives. Due to its popularity, social networking has given rise to a number of issues, including the potential for dangerous content to spread by tricking people into believing they are someone they are not. This circumstance has the potential to cause significant harm to society in the actual world. In our study, we offer a classification technique for identifying Twitter bogus accounts. Our dataset was pre-processed using the Entropy Minimization Discretion (EMD) method on numerical features explained.

4. OUTPUT SCREENS

Home Page

View profile page

View Remote Users

Output

ISSN 2347–3657

Volume 12, Issue 3, 2024

5. CONCLUSION

We presented natural language processing and machine learning methods in this article. We may readily identify phony accounts on social media by using these methods. Finding the false profiles was the focus of this research, which used the Facebook Data set. In order to examine the dataset, natural language processing (NLP) pre-processing methods are used, and machine learning algorithms like SVM and Naïve Bayes are employed for profile classification. In this study, we show that these learning techniques may increase the detection rate.

6. REFERENCE

[1] Social Networks Analysis and Mining (ASONAM) 2018 Aug 28 (pp. 1191-1198). IEEE.

[2] Pakaya FN, Ibrohim MO, Budi I. Malicious Gheewala S, Patel R. ML based Twitter Spam account detection: a review. In 2018 Second International Conference on Computing Methodologies and Communication (ICCMC) 2018 Feb 15 (pp. 79-84). IEEE.

[3] Kaliyar RK. Fake news detection using a deep neural network. In 2018 4th International Conference on Computing Communication and Automation (ICCCA) 2018 Dec 14 (pp. 1-7). IEEE.

[4] Erşahin B, Aktaş Ö, Kılınç D, Akyol C. Twitter fake account detection. In2017 International Conference on Computer Science and Engineering (UBMK) 2017 Oct 5 (pp. 388-392). IEEE.

[5] Gupta A, Kaushal R. Towards detecting fake user accounts in Facebook. In2017 ISEA Asia Security and Privacy (ISEASP) 2017 (pp. 1-6). IEEE.

[6] Alom Z, Carminati B, Ferrari E. Detecting spam accounts on Twitter. In 2018 IEEE/ACM International Conference on Advances in Account Detection on Twitter Based on Tweet Account Features using Machine Learning. In 2019 Fourth International Conference on Informatics and Computing (ICIC) 2019 Oct 16 (pp. 1-5). IEEE.

ISSN 2347–3657

Volume 12, Issue 3, 2024

[7] Jardaneh G, Abdelhaq H, Buzz M, Johnson D. Classifying Arabic tweets based on credibility using content and user features. In2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT) 2019 Apr 9 (pp. 596-601). IEEE.

[8] Harjule P, Sharma A, Chouhan S, Joshi S. Reliability of News. In2020 3rd International Conference on Emerging Technologies in Computer Engineering: ML and Internet of Things (ICETCE) 2020 Feb 7 (pp. 165-170). IEEE.

[9] Dr.C K Gomathy, Article: A Study on the recent Advancements in Online Surveying , International Journal of Emerging technologies and Innovative Research (JETIR) Volume 5 Issue 11 | ISSN : 2349-5162, P.No:327-331, Nov-2018

[10] B. Erçahin, Ö. Akta³, D. Kilinç, and C. Akyol, ``Twitter fake accountdetection," in Proc. Int. Conf. Comput. Sci. Eng. (UBMK), Oct. 2017,pp. 388_392.