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 15, Issue 3, 2024

https://doi.org/10.62646/ijitce.2024.v15.i3.pp853-857

FAKE PROFILES IDENTIFICATION IN ONLINE SOCIAL NETWORKS USING MACHINE LEARNING AND NLP

[1] Mrs.Bethapudi ArunaSri,[2]M. Rajeshwari,[3] P. Yashaswini,[4]P. Joshna
[1]Assistant Professor,[234]Students
Department Of CSE
Malla Reddy Engineering College for Women

ABSTRACT_ Social media is an integral part of most people's everyday routines nowadays. All hours of the day and night, countless people join social media sites to connect with others and share information. Although social media sites have many positive uses, they also pose risks to users' personal information. In order to find out who is spreading threats on social media, we need to classify user profiles. Classification allows us to differentiate between authentic and fake social media profiles. The detection of fake social media accounts has long made use of a number of classification schemes. On the other hand, we need to improve the precision with which social media platforms identify fake profiles. Using methods from machine learning and natural language processing (NLP), this study aims to improve the false profile identification accuracy rate. Support Vector Machine (SVM) and the Naïve Bayes technique are both applicable.

1.INTRODUCTION

Nowadays, social networking has become a well-known online pastime, drawing hundreds of thousands of users who log on for billions of minutes each year. The range of online social network (OSN) services includes social interaction-focused platforms like Facebook and MySpace, comprehension-focused platforms like Twitter and Google Buzz, and social interaction features integrated into current systems like flicker. On the other hand, strengthening security issues and safeguarding OSN privacy continue to be recognized as a primary bottleneck and purpose.

One of a kind men and women communicate one-of-a-kind volumes of their private knowledge via social networks (SNs). Because our personal knowledge is either fully or partially disclosed to the public, we are prime candidates for some kinds of attacks, the worst of which may be identity theft. Identity theft occurs when someone appropriates a character's knowledge for their own gain or agenda. In previous times, millions of individuals worldwide were impacted by online identity theft, making it a major issue. Identity theft victims may face special punishments, such as losing money or time, being sent to reformatory, having their reputation harmed, or experiencing problems in their relationships with friends and family. Nowadays, the great majority of SNs have extremely shaky privacy and security standards and do longer verify the debts of regular customers. In actuality, the majority of SN apps have low privacy settings by default, which has made SNs an ideal venue for fraud and abuse. Social networking platforms have made identity theft and impersonation attempts easier for

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 15, Issue 3, 2024
https://doi.org/10.62646/ijitce.2024.v15.i3.pp853-857

both skilled and unsuspecting attackers. To exacerbate matters, users must provide accurate information in order to create an account on social networking websites. Simple oversight of what users post online might result in disastrous losses, never mind if these bills were compromised.

Online networks will also include static or dynamic profile data. Static knowledge refers to the information that may be provided by the user during the profile construction process, while dynamic knowledge refers to the fine print that is narrated by the network system. A person's realtime behaviors and location within the network are examples of dynamic knowledge, while demographic information and interests about the individual make up static knowledge. Most of the research being done now relies on both static and dynamic data. This isn't applicable to many social networks, however, since only a small percentage of static profiles are seen and dynamic profiles are often hidden from the human network. A unique researcher has presented a number of methods to identify fraudulent information and false identities in online social networks. Every procedure has strengths and drawbacks of its own.

Social networking issues, including those related to privacy, cyberbullying, abuse, trolling, and many more. are several occasions when fraudulent accounts on social networking platforms are used. Unspecific profiles are known as false profiles. These are the fictitious profiles of men and women. False Facebook accounts often engage in harmful and undesired behavior, which disrupts users of social media networks. False profiles are made by people for social engineering, online impersonation to harm a person, and campaigning and supporting a persona or group of people. Facebook has its own security system in place to protect user credentials from phishing, spam, and other threats. It's also often referred to as the Facebook Immune System (FIS).
The FIS hasn't been prepared to see more and more phony Facebook accounts created by users.

2.LITERATURE SURVEY

2.1 Dr. S. Kannan, Vairaprakash Gurusamy, “Preprocessing Techniques for Text Mining”, 05 March 2015.

Preprocessing is a crucial stage in information retrieval (IR), text mining, and natural language processing (NLP). Data preprocessing is used in the field of text mining to extract knowledge and meaningful, non-trivial information from unstructured text data. Selecting which documents in a collection should be retrieved to meet a user's information request is the essence of information retrieval (IR). A query or profile, which includes one or more search phrases and extra metadata like word weight, represents the user's information requirement.

2.2 Shalinda Adikari and Kaushik Dutta, Identifying Fake Profiles in LinkedIn, PACIS 2014 Proceedings, AISeL

Social networks are already a commonplace tool in our lives, and their target audiences vary widely. People in professional vocations tend to like LinkedIn the most among them. Social networks are growing quickly, and

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 15, Issue 3, 2024
https://doi.org/10.62646/ijitce.2024.v15.i3.pp853-857

individuals are using them for immoral and unlawful purposes. The creation of a false profile has an adversarial impact that is hard to pinpoint without adequate investigation. The features and social network connections of the user's social profile have been taken into consideration in the majority of the current solutions that have been theoretically and practically created to resolve this dispute. However, LinkedIn's privacy standards severely limit what is publicly accessible about users' profiles, including behavioral observations. Due to LinkedIn's low amount of publicly accessible profile data, it is not possible to use the current methods for identifying fraudulent profiles. Thus, it is necessary to carry out focused study on methods for detecting phony LinkedIn profiles. In this study, we determine the smallest amount of profile data required to detect fraudulent profiles on LinkedIn, as well as the best data mining technique for the job. We show that our method can detect the phony profile with 84% accuracy and just 2.44% false negative with limited profile data. This is similar to the findings of other current methods based on a bigger data set and more profile details.

2.3 Z. Halim, M. Gul, N. ul Hassan, R. Baig, S. Rehman, and F. Naz,"Malicious users' circle detection in social network based on spatiotemporal co-occurrence," in Computer Networks and Information Technology(ICCNIT),2011 International Conference on, July, pp. 35–390.

Over the last decade, there has been a marked improvement in the way users see online social networks. It is becoming more popular with spammers as well. Tracking spammers on a large scale is no easy feat. Because they interact covertly, spammers are immune to detection by even the most basic social network graph analysis. By classifying people according to their spatiotemporal co-occurrence, or communication frequency, we are able to determine the group of people involved in the detrimental communications. Grouping persons based on their spatiotemporal co-occurrence allows us to identify the users engaging in hostile messages in this study.

RELATED WORK

The prize Chai et al. received for their article is evidence of their inspiration and expertise. The results obtained from the user testing of the prototype technique are noteworthy, even if it has used the most efficient regular systems in natural language processing and human-computer interaction. Through the comparison of this basic prototype method with a fully implemented menu process, they have found that users—primarily novice users—strongly prefer the common language dialog-based method. They have also discovered that, in an e-commerce setting, dialog administration skill is more crucial than the ability to handle intricate phrases in ordinary English. Additionally, menu-pushed navigation and natural language dialog-based navigation should be cleverly integrated to satisfy each user's unique needs in order to enable simple access to information on e-commerce websites. They have completed the construction of a new version of the strategy that offers significant improvements in information management, dialog administration, and language processing. They felt that informal interfaces with ordinary language provide strong, customized alternatives to

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 15, Issue 3, 2024
https://doi.org/10.62646/ijitce.2024.v15.i3.pp853-857

traditional menu-driven or search-based website interfaces.

2 .PROPOSED SYSTEM

• To identify phoney social media accounts, we presented a method that combines machine learning with natural language processing in this research. Furthermore, to enhance the precision of false profile detection, the SVM classifier and naïve bayes method are being included. A Support Vector Machine (SVM) finds the one-of-a-kind hyperplane that separates all data features of one kind from those of another type in order to classify the data. When using a support vector machine (SVM) method, the best hyperplane to choose is the one with the longest line linking the two classes. One way support vector machines (SVMs) categorise data is by using the exceptional hyperplane, which separates all knowledge components of two classes. The assistance vectors are the data points that are geographically nearest to the separation hyperplane.

Using a set of predefined criteria, the Naive Bayes algorithm may predict which group an object is most likely to be a member of. Simply said, it is a probabilistic classifier. One reason the Naive Bayes method is called "naive" is because it presumes that the presence of a distinct feature is unrelated to the frequency of other characteristics. Say, for the sake of argument, that we are interested in detecting phoney accounts using the following criteria: time, geolocation, language, and the publishing date of posts. Even if they depend on each other or other factors, I still think all of these things make the phoney profile more likely to exist.

3.1 IMPLEMENTTAION

Service Provider

To access this module, the Service Provider has to provide a valid username and password. After he logs in, he'll be able to perform things like train and evaluate electrical data sets, Check the Reliability of Power Databases using a Bar Chart, Check the Power Datasets' Accuracy in the Results You may see all the distant users, the predicted data sets, the theft outcomes, the expected sort of theft, and the ratio of theft to type. Access Any User From Anywhere.

View and Authorize Users

A complete roster of all users registered for this module is seen by the administrator. Users' names, email addresses, and physical addresses are viewable to the administrator, who may also authorise users.

Remote User

In this module, you will find n users. The user must register before they may begin any activity. Upon registration, the user's details are stored in the database. After signing up, he'll need to log in using the credentials that were given to him. Upon completion, the user is able to access many features, including the ability to register and log in, make predictions about the kind of energy theft, and see their profile.

4. LEARNING ALGORITHMS

Support Vector Machine (SVM) and naïve Bayes methods are used in this suggested system.

Support Vector Machine (SVM)

A support vector machine (SVM) sorts data into categories by locating the unique hyperplane that divides all data points into

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 15, Issue 3, 2024
https://doi.org/10.62646/ijitce.2024.v15.i3.pp853-857

two groups, one for each categorisation. The hyperplane with the longest distance between the two categories is the optimal one for a support vector machine (SVM) approach. A support vector machine (SVM) finds the exceptional hyperplane that divides all knowledge aspects of one class from those of the other class in order to classify data. The data points that are geographically closest to the separation hyperplane are known as the help vectors.

5.CONCLUSION

In this research, we combined natural language processing methods with machine learning algorithms. These methods make it simple to identify fraudulent personas on social media platforms. To detect the phony profiles, we used the Faceboo dataset in this research. The dataset is analyzed using NLP pre-processing methods, and the profiles are classified using machine learning algorithms like SVM and Naïve Bayes. In this research, the detection accuracy rate is increased using these learning techniques.

REFRENCES

[1] Michael Fire et al. (2012). "Strangers intrusion detection-detecting spammers and fake profiles in social networks based on topology anomalies." Human Journal 1(1): 26-39.Günther, F. and S. Fritsch (2010). "neuralnet: Training of neural networks." The R Journal 2(1): 30-38

[2] Dr. S. Kannan, Vairaprakash Gurusamy, “Preprocessing Techniques for Text Mining”, 05 March 2015.

[3] Shalinda Adikari and Kaushik Dutta, Identifying Fake Profiles in LinkedIn, PACIS 2014 Proceedings, AISeL

[4] Z. Halim, M. Gul, N. ul Hassan, R. Baig, S. Rehman, and F. Naz,“Malicious users’ circle detection in social network based on spatiotemporal co-occurrence,” in Computer Networks and Information Technology (ICCNIT),2011 International Conference on, July, pp. 35–390.

[5] Liu Y, Gummadi K, Krishnamurthy B, Mislove A,” Analyzing Facebook privacy settings: User expectations vs. reality”, in: Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference,ACM,pp.61–70.

[6] Mahmood S, Desmedt Y,” Poster: preliminary analysis of google?’s privacy. In: Proceedings of the 18th ACM conference on computer and communications security”, ACM 2011, pp.809–812.

[7] Stein T, Chen E, Mangla K,” Facebook immune system. In: Proceedings of the 4th workshop on social network systems”, ACM 2011, pp

[8] Saeed Abu-Nimeh, T. M. Chen, and O. Alzubi, “Malicious and Spam Posts in Online Social Networks,” Computer, vol.44, no.9, IEEE2011, pp.23– 28.

[9] J. Jiang, C. Wilson, X. Wang, P. Huang, W. Sha, Y. Dai, B. Zhao, Understanding latent interactions in online social networks, in: Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement, ACM, 2010, pp. 369–382

[10] Kazienko, P. and K. Musiał (2006). Social capital in online social networks. Knowledge-Based Intelligent Information and Engineering Systems, Springer