International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 12 Dec 2021, pp: 1089-1094 www.ijaem.net ISSN: 2395-5252

A Surevey on Machine Learning Algorithms for Identifying Fake Identities

[1]Versha Dadhore, [2]Prof. Nidhi Dubey
[1 2] *Department of Computer Science & Engineering*
Bhopal Institute of Technology & Science, Bhopal

Submitted: 05-12-2021 Revised: 17-12-2021 Accepted: 20-12-2021

ABSTRACT--False identities play an important role in advanced threats which are also involved in other malicious activities. The present article focuses on the literature review of the state-of-the-art research aimed at detecting fake profiles in social media. The approaches to detecting fake social media accounts can be classified into the approaches aimed on analysing individual accounts, and the approaches capturing the coordinated activities spanning a large group of accounts. The paper sheds light on the role of fake identities in advanced persistent threats and covers the mentioned approaches of detecting fake social media accounts. Thus, the analysis for detecting fake accounts will be done by applying suitable algorithm.

Keywords: *Social Network Analysis, Social Media, Fake Profiles, False Identities.*

I. INTRODUCTION

Identity is an object attached to a human being, separate from him or her. A typical example is the name of a person. Another example is a passport that contains the name, birth date and place of the person, nationality, digitally captured fingerprints and a digitally stored and a photograph of the person. A third example is a private and public key adhering to a Public Key Infrastructure. In general, identity should be unique in the sense that each identifying object must only refer to at most one person. A similar individual may at present have a few personalities, similar to an international ID and a couple of keys above, or a government disability number. The genuine character is confirmed by experts of some country state.

False identities play an important role in advanced persisted threats (APT), i.e. coordinated, lasting, complex efforts at compromising targets in governmental, non-governmental, and commercial organizations. False identities are also often involved in other malicious activities, like spamming, artificially inflating the number of users in an application to promote it, etc. A typical scenario for using false identities is using social media platforms to impersonate someone or create a fake identity to establish trust with the target, which is then exploited:

♣for gathering further information for a spear phishing attack,

♣mounting a spear phishing attack, or

♣for directly interacting to get the information of interest. In the sequel we consider originally authentic, but later compromised accounts as false accounts. We also call false such accounts that contain personal information, which does not belong to the person who created this account. If the account contains, invented personal details it is called a faked account Items that are taken as identifiers must be certified by the authorities of a country of issue, recognized inside this country, and beyond its bounds with a mutual agreement with other

A cutting edge visa is an ordinary case of this. Experts ensure that the image, fingerprints, name, birthdate and so forth have a place with a similar individual, for example ensure the item connection. At an online networking webpage a client is normally recognized by a profile. It normally contains an image and name, perhaps a location and birth date. The destinations don't, be that as it may, thoroughly watch that the individual with the character implied in the profile truly made and controls the profile. On the off chance that this isn't the situation, someone is utilizing another person's character. This is called false personality. One can likewise make profiles that can utilize unreservedly designed names and other data that can't be joined to any genuine individual in any nation. For this situation the character is known as a faked personality. Such a profile can at present contain an image of a genuine individual, picked for example haphazardly from the Internet. False characters assume a significant job in cutting edge endured

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 12 Dec 2021, pp: 1089-1094 www.ijaem.net ISSN: 2395-5252

dangers (APT), for example facilitated, enduring, complex endeavors at trading off focuses in administrative, non-legislative, and business associations. False characters are likewise regularly associated with different malignant exercises, such as spamming, misleadingly expanding the quantity of clients in an application to advance it, and so forth. A normal situation for utilizing false personalities is utilizing web based life stages to imitate somebody or make a phony character to set up trust with the objective, which is then abused:

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves. The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide.

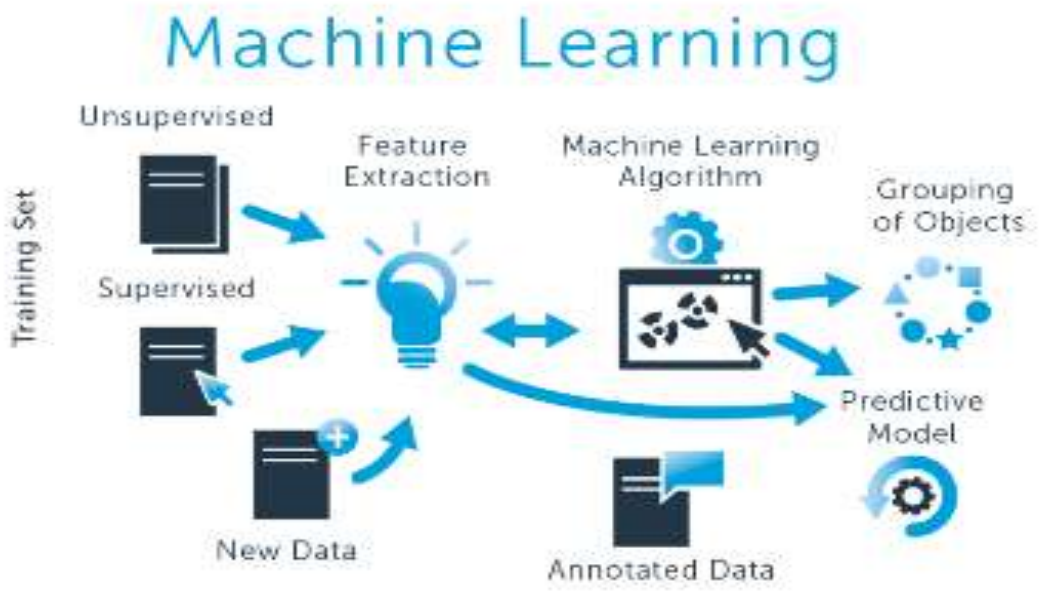

Figure 1: Machine Learning.

 For social affair additional data for a lance phishing attack,
 Mounting a lance phishing attack, or
For legitimately associating to get the data of intrigue. In the continuation we consider initially real, however later traded off records as false records. We additionally consider false such records that contain individual data, which does not have a place with the individual who made this record. In the event that the record contains, imagined individual subtleties it is known as a faked record Items that are taken as identifiers must be confirmed by the experts of a nation of issue, perceived inside this nation, and past its limits with a common concurrence with other.

➢ Some machine learning methods

Machine learning algorithms are often categorized as supervised or unsupervised.

- Supervised machine learning algorithms can apply what has been learned in the past to new data using labeled examples to predict future events. Starting from the analysis of a known training dataset, the learning algorithm produces an inferred function to make predictions about the output values. The system is able to provide targets for any new input after sufficient training. The learning algorithm can also compare its output with the correct, intended output and find errors in order to modify the model accordingly.
- In contrast, unsupervised machine learning algorithms are used when the information used to train is neither classified nor labeled. Unsupervised learning studies how systems can infer a function to describe a hidden structure from unlabeled data. The system doesn't figure out the right output, but it explores the data and can draw inferences from datasets to describe hidden structures from unlabeled data.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 12 Dec 2021, pp: 1089-1094 www.ijaem.net ISSN: 2395-5252

- Semi-supervised machine learning algorithms fall somewhere in between supervised and unsupervised learning, since they use both labeled and unlabeled data for training – typically a small amount of labeled data and a large amount of unlabeled data. The systems that use this method are able to considerably improve learning accuracy. Usually, semi-supervised learning is chosen when the acquired labeled data requires skilled and relevant resources in order to train it / learn from it. Otherwise, acquiringunlabeled data generally doesn't require additional resources.
- Reinforcement machine learning algorithms is a learning method that interacts with its environment by producing actions and discovers errors or rewards. Trial and error search and delayed reward are the most relevant characteristics of reinforcement learning. This method allows machines and software agents to automatically determine the ideal behavior within a specific context in order to maximize its performance. Simple reward feedback is required for the agent to learn which action is best; this is known as the reinforcement signal.

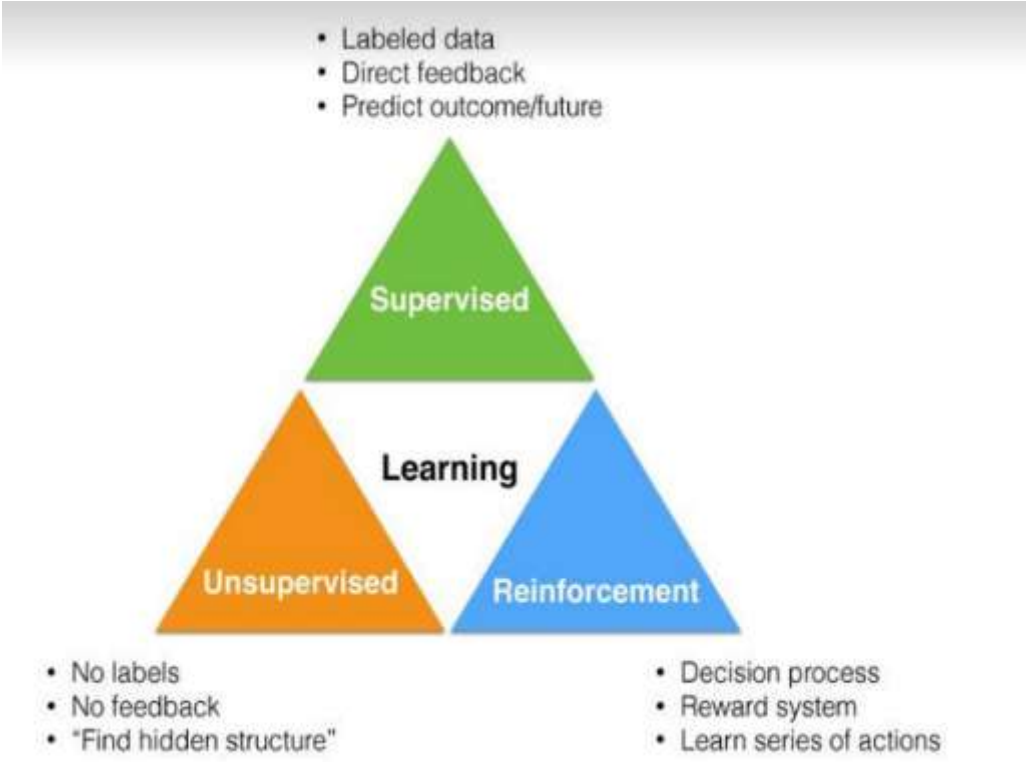

Figure 2: Machine Learning Category.

In the above figure 2 the category of the machine learning is shown.

II. LITERATURE REVIEW

A number of fake account detection approaches rely on the analysis of individual social network profiles, with the aim of identifying the characteristics or a combination thereof that help in distinguishing the legitimate and the fake accounts. Specifically, various features are extracted from the profiles and posts, and then machine learning algorithms are used in order to build a classifier capable of detecting fake accounts (Table 1).

1. For example, the paper **Nazir et al. (2010)** depicts recognizing and portraying apparition profiles in online social gaming applications. The paper investigations a Facebook application, the web based diversion "Warriors club", known to give motivating forces and gaming preferred standpoint to those clients who welcome their looks into the amusement. The Authors contend that by giving such motivators the amusement inspires its players to make counterfeit profiles. By bringing those phony profiles into diversion, the client would expand motivator esteem for him/herself. At first, the Authors remove 13 highlights for each amusement client, and afterward perform grouping

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 12 Dec 2021, pp: 1089-1094 www.ijaem.net ISSN: 2395-5252

utilizing bolster vector machines (SVMs). The paper presumes that these strategies don't recommend any undeniable discriminants among genuine and counterfeit clients [6].

2. **Adikari and Dutta (2014)** [7] depict ID of phony profiles in LinkedIn. The paper demonstrates that phony profiles can be distinguished with 84% exactness and 2.44% false negative, utilizing constrained profile information as information. Strategies, for example, neural systems, SVMs, and foremost segment examination are connected. Among others, highlights, for example, number of dialects spoken, instruction, aptitudes, suggestions, interests, and grants are utilized. Attributes of profiles, known to be phony, posted on exceptional sites are utilized as a ground truth.

3. **Chu et al. (2010)** [8] go for separating Twitter accounts worked by human, bots, or cyborgs (i.e., bots and people working in show). As a piece of the discovery issue plan, the recognition of spamming accounts is acknowledged with the assistance of an Orthogonal Sparse Bigram (OSB) content classifier that utilizes sets of words as highlights. Went with other recognizing parts surveying the normality of tweets and some record properties, for example, the recurrence and sorts of URLs and the utilization of APIs, the framework had the capacity to precisely recognize the bots and the human-worked accounts.

4. Distinguishing spamming accounts in Twitter just as in MySpace, was additionally the target of the investigation by Lee et al. (2010) [9]. As contrasted and the examination by Chu et al., the arrangement of highlights here was extended to cover additionally the number and sort of associations. Various classifiers accessible in Weka AI suite were attempted, and the Decorate meta classifier was found to give the best order exactness.

5. Notwithstanding, or as opposed to breaking down the individual profiles, another surge of methodologies depend on chart based highlights while recognizing the phony and authentic records. For example, **Stringhini et al. (2010)** [10] portray strategies for spam discovery in Facebook and Twitter. The Authors made 900 honeypot profiles in informal communities, and performed constant accumulation of approaching messages and companion demands for a year. Client information of the individuals who played out these solicitations were gathered and broke down, after which about 16K spam accounts were identified. Authors further explored the utilization of AI for further location of spamming profiles. Over the highlights utilized in the examinations over, the Authors were additionally utilizing the message comparability, the nearness of examples behind the hunt of companions to include, and the proportion of companion solicitations, and afterward utilized Random Forest as a classifier.

Krombholz et al. (2015) [11] proposes classification of social engineering attacks into physical methods (such as dumpster diving), social approaches (relying on socio-psychological techniques), reverse social engineering (attacker attempts to make victim believe that she is a trustworthy entity, and the goal is to make the victim approach attacker e.g. for help), technical approaches, and socio-technical approaches (combining approaches above). Kontaxis et al. (2011) [12] describe prototype of the software which aims at finding whether profile of particular user was cloned from one online social network into another by comparing characteristics of the profiles having similar characteristics among several online social networks.

Krombholz et al. (2012) [13] propose the raising of users' awareness as the most efficient countermeasure against social media identity theft, and describes the methods for it. Authors perform focus groups research, and suggest that the users are mostly unaware of fake profiles occurrence and its consequences. Jiang et al. (2016) [14] surveyed more than 100 advanced techniques for detecting suspicious behaviors that have existed over the past 10 years and presented several experimentally successful detection techniques (i.e. CopyCatch, which was described in (Beutel et al., 2013) [15]).

Table 1: Profile-based methods for detecting fake social media accounts.

Reference	Ground truth	Detection method
Adikari 2015	Known fake LinkedIn profiles, posted on special web sites	Number of languages spoken, education, skills, recommendations, interests, awards, etc. are used as features to

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 12 Dec 2021, pp: 1089-1094 www.ijaem.net ISSN: 2395-5252

		train neural networks, SVMs, and principal component analysis.
Chu et al. 2010	Manually labelled 3000x2 Twitter profiles as human, bots, or cyborgs.	Manually labelled 3000x2 Twitter profiles as human, bots, or cyborgs.
Lee et al. 2010	Spam accounts registered by honeypots: 1500 in MySpace and 500 in Twitter	Over 60 classifiers available in Weka are tried. Features include: i) demographics, ii) content and iii) frequency of content generation, iv) number and type of connections. The Decorate meta-classifier provided the best results.
Stringhini et al. 2010	Spam accounts registered by honeypots: 173 spam accounts in Facebook and 361 in Twitter	Random forest was constructed based on the following features: ratio of accepted friend requests, URL ratio, message similarity, regularity in the choice of friends, messages sent, and number of friends.
Yang et al. 2011a	Spam Twitter accounts defined as the accounts containing malicious URLs: 2060 spam accounts	Graph based features (local clustering coefficient, between centrality, and bi-directional links ratio), neighbor-based features (e.g., average neighbors' followers), automation-based features (API ratio, API URL ratio and API Tweet similarity), and timing-based features were used to construct different classifiers.

In the comparison table 1 above, some existing recent algorithms are discussed, their advantages, disadvantages, limitation and further extension is discussed in the given table.

III. CONCLUSION

False identities in the form of compromised or fake email accounts, accounts in social media, fake or cracked websites, fake domain names, and malicious Tor nodes, are heavily used in APT attacks, especially in their initial phases, and in other malicious activities. Using these fake identities, the attacker(s) aim at establishing trust with the target and at crafting and mounting a spear phishing or another attack. Based

on paper evidence, information gathering for a spear phishing attack heavily relies on the use of social media and fake accounts therein. It is therefore important to detect, as early as possible, the presence of a fake social media account. A number of recent paper works have focused on detecting such fake accounts, either by analysing the characteristics of individual profiles and their connections, or – in case of coordinated activities, by multiple fake social media accounts, such as in the case of crowd turfing – by analysing the commonality of these activities, too.

REFERENCES

[1]. S. Gu rajala, J. S. White, B. Hudson, B. R. Voter, and J. N. Matthews, "Profile characteristics of fake twitter accounts," Big Data & Society, vol. 3, no. 2, p. 2053951716674236, 2016.

[2]. C. Xiao, D. M. Freeman, and T. Hwa, "Detecting clusters of fake accounts in o nline social networks," in Proceedings of the 8th ACM Workshop on Artificial Intelligence and Security. ACM, Conference Proceedings, pp. 91 – 101.

[3]. S. Mainwaring, We first: How brands and consumers use social media to build a better world. Macmillan, 2011.

[4]. V. Subrahmanian, A. Azaria, S. Durst, V. Kagan, A. Galstyan, K. Lerman, L. Zhu, E. Ferrara, A. Flammini, and F. Menczer, "The darpa twitter bot challenge," arXiv preprint arXiv:1601.05140, 2016.

[5]. Y. Li, O. Martinez, X. Chen, Y. Li, and J. E. Hopcroft, "In a w orld that counts: Clustering and detecting fake social engagement at scale," in Proceedings of the 25th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, Conference Proceedings, pp. 111 – 120.

[6]. Nazir, A., Raza, S., Chuah, C.-N., Schipper, B., 2010.Ghostbusting Facebook: Detecting and Characterizing Phantom Profiles in Online Social Gaming Applications, in: Proceedings of the 3rd Wonference on Online Social Networks, WOSN'10. USENIX Association, Berkeley, CA, USA, pp. 1–1.

[7]. Adikari, S., Dutta, K., 2014. Identifying Fake Profiles in Linkedin, in: PACIS 2014 Proceedings.Presented at the Pacific Asia Conference on Information Systems.

[8]. Chu, Z., Gianvecchio, S., Wang, H., Jajodia, S., 2010. Who is Tweeting on Twitter: Human, Bot, or Cyborg?, in: Proceedings of the 26th Annual Computer Security Applications Conference, ACSAC '10. ACM, New York, NY, USA, pp. 21–30. doi:10.1145/ 1920261.1920265.

[9]. Fayazi, A., Lee, K., Caverlee, J., Squicciarini, A., 2015. Uncovering Crowdsourced Manipulation of Online Reviews, in: Proceedings of the 38th International ACM SIGIR Conference on Paper and Development in Information Retrieval, SIGIR '15. ACM, New York, NY, USA, pp. 233–242. doi:10.1145/2766462.2767742.

[10]. Egele, M., Stringhini, G., Kruegel, C., Vigna, G., 2015. Towards Detecting Compromised Accounts on Social Networks. IEEE Trans. Dependable Secure Comput.PP, 1–1. doi:10.1109/TDSC.2015.2479616.

[11]. Krombholz, K., Hobel, H., Huber, M., Weippl, E., 2015. Advanced Social Engineering Attacks. J InfSecurAppl 22, 113–122. doi:10.1016/j.jisa.2014.09.005.

[12]. Kontaxis, G., Polakis, I., Ioannidis, S., Markatos, E.P., 2011. Detecting social network profile cloning, in: 2011 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops). Presented at the 2011 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops), pp. 295–300. doi:10.1109/PERCOMW.2011.5766886.

[13]. Krombholz, K., Merkl, D., Weippl, E., 2012. Fake identities in social media: A case study on the sustainability of the Facebook business model. J. Serv. Sci. Res. 4, 175–212. doi:10.1007/s12927-012-0008-z.

[14]. Jiang, M., Cui, P., Faloutsos, C., 2016. Suspicious Behavior Detection: Current Trends and Future Directions. IEEE Intell. Syst. 31, 31–39. doi:10.1109/MIS.2016.5.

[15]. Beutel, A., Xu, W., Guruswami, V., Palow, C., Faloutsos, C., 2013. CopyCatch: Stopping Group Attacks by Spotting Lockstep Behavior in Social Networks, in: Proceedings of the 22Nd International Conference on World Wide Web, WWW '13. ACM, New York, NY, USA, pp. 119–130. doi:10.1145/2488388.2488400.