Literature review

"Machine Learning Techniques for Online Identity Fraud Mitigation"

Name: Viraj Sunil Oke

Student ID: 300251185

Date: 26/10/2021

University of Ottawa

Abstract- **With the increased usage of social media, the risk of identity fraud and associated cyber security threats has flourished over a period of time. Identity fraud can be carried out mainly by two actors, humans and bots. Both the actors are capable of falsifying social media accounts and creating duplicate profiles. One of the most common methods used by the actors is, creating multiple/duplicate accounts. The actors such as bot programs can run on AI algorithms that can mimic the human traits for creating duplicate accounts. On the other hand, human actors creating multiple accounts with the same identity to falsify the number of social media followers is a common cyber security concern. Differentiating between these two actors can be crucial for the efficient mitigation of this cyber security threat. When it comes to Identity fraud mitigation, machine learning techniques such as classification, clustering, and regression techniques can be used with an integration of various data collection and sampling methods for effective detection of fake profiles online. In this review, the discussed research is about a number of machine learning techniques for the detection of identity deception on social media and its mitigation.**

1. Introduction

Social media platforms have evolved over the years in terms of the scope of functionalities and the way it enables people to connect around the globe. However, with an increase in the number of social media users, a concern of cyber security threats has increased exponentially (Walt et al., 2018). Cyber security concerns can be classified into a wide spectrum of threats, but most of the threats related to social media originate from fake identities created by humans or bots to exploit other users for personal gains (Van Der Walt & Eloff, 2018; Walt et al., 2018). Identity deception is a major concern when it comes to social media platforms. The reason is, users have their personal sensitive data exposed on the internet hence increasing the chances of fake account creation (Borkar & Sharma, 2020). The process of falsifying accounts can be carried out by not just humans but also by the bot systems which are programmed to create fake accounts automatically using human traits (Belokurov et al., 2021). "A Fake account is a representation of an individual or organization who pretends to be someone else", Suganya et al (2021). A lot of social media sites including Facebook and Twitter are affected by identity deception. Some studies have shown that 51.8% of the web traffic was observed to be created by bots and around 13.5 million fake accounts were reported by Twitter. Moreover, the numbers are incrementing exponentially on a year-to-year basis (Gilani et al., 2017). To overcome this problem, various ML approaches can be used in order to mitigate ID fraud on big data efficiently.

2. Methods

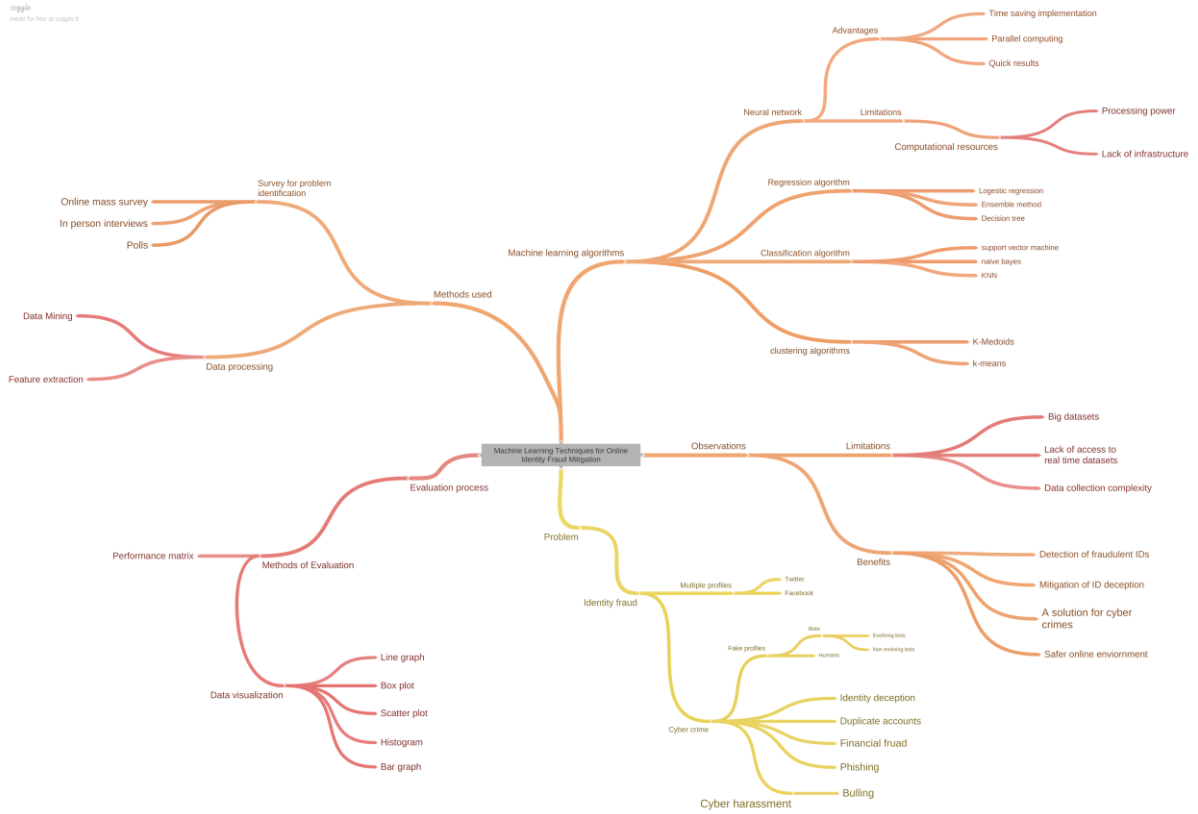
The objective of this review is to detect and mitigate identity fraud on social media which is a rapidly increasing cyber security threat. The cyber threat of ID fraud exists on many social media platforms such as Facebook, Twitter, LinkedIn, Google+ (Khaled & El-tazi, 2018; Er et al, 2017, Xiao et al, 2015).

The review tries to answer the question of how to bridge the gaps between the existing methodologies using various machine learning techniques and data processing methods which can yield desired results for all the social media platforms, unlike the previous studies which only focused on specific platforms such as Facebook or Twitter.

Future scope: The existing research is limited to detecting fake profiles specifically on social media platforms. However, the research can be extended to solve ID frauds on other online platforms as well. For example, net banking, reservation portals, academic portals, etc.

The sources used for the review are Scopus, Web of Science, IEEE Xplore, and other peer-reviewed sources. The methods and algorithms discussed in the review are Support Vector Machine (SVM), Naive Bayes, and many more. The clustering algorithms discussed are K-Means and K-Medoid. For problem Identification, the results of various surveys have been analyzed. For data processing, the concepts such as data mining and feature selection have been discussed.

2.1. Organization of the review:



3. Literature Review

3.1. Problem

Fake identities online:

Cyber threats associated with fake profiles online are diverse. From spamming, privacy breaches, phishing, duplicate accounts to terrorism. It all originates from fake profiles. Either created by humans or bots (Van Der Walt & Eloff, 2018). When it comes to detecting fake profiles online in a Bigdata environment a lot of factors comes into play such as, how well the human traits have been impersonated by the bots, how well the human actor has replicated the victim profile for identity deception (Van Der Walt & Eloff, 2018; Narayanan, 2021). It is important to consider duplicate or multiple accounts while detecting ID frauds due to their sheer volume and the impact they can have on a large scale. A study has shown that Facebook had over 10 Billion accounts in 2019. However, there were just 2 billion users on Facebook in 2019+ which clearly implies that there is a problem of identity deception on social media (Wang et al., 2019) and these multiple accounts could be hard to detect if they are well-crafted fake profiles with convincing human traits (Wang et al., 2019). When multiple accounts are created by bots, the process of detecting them becomes difficult if the bots are "evolving bots" as they can evolve their human traits with time (Van Der Walt & Eloff, 2018; Gilani et al, 2017). Bot programs with the kind of nature they exhibit are easier to detect as opposed to human actors cause they tend to follow a pattern while mimicking human traits (Belokurov et al., 2021; Van Der Walt & Eloff, 2018) as a solution to this problem "Captcha" is seen to be an effective means of mitigating bot intervention on the internet (Xiao et al., 2015).

The problem identification process consists of surveys and polls over social media (Narayanan, 2021). These methods can make the data collection process more reliable as the data is directly collected from the end-user. Apart from surveys and human intervention for investigation, human actors and bots can be detected using machine learning techniques for a large set of datasets like Twitter, where the algorithms can detect the bots with 87.84% accuracy (Suganya et al., 2021). However, the first step of developing machine learning models is data collection and pre-processing, which will be discussed in the next sub-section.

3.2. Methods used for data collection and feature extraction

a) Data mining:

The process of data collection from social media platforms involves extracting data from Bigdata. To extract the datasets with relevant features, technologies such as Hadoop, DBMS, and different APIs are used (Van Der Walt & Eloff, 2018). The extracted features can be grouped into various categories such as individual profiles, groups of people, and the links between different nodes. Graph theory can be used to find the associations between those nodes (Wang et al., 2019). To reduce the processing load on the system, techniques like

Principle component analysis PCA can be used to reduce feature dimensions which makes the process of extracting relevant data easier (Khaled & El-tazi, 2018).

b) *Feature extraction:*

Feature extraction is one of the most important steps for building a machine learning solution. While working on a big dataset, it should be taken into consideration that the dataset is a large pool of features that can be relevant or irrelevant for the machine learning model. Choosing irrelevant features might result in higher entropy value (Walt et al., 2018). A few examples of choosing relevant features for social media platforms are Name, Gender, Follower count, Friends count, Creation date, Location, and Language (Singh et al, 2018; Narayanan, 2021; Walt et al., 2018). After using a subset of these features to develop an Identity deception detection model, an F1 score of 86.24% was recorded (Walt et al., 2018). A platform-oriented feature selection for Twitter can consist of features such as follower count, friends count, retweets, user favorites, user tweets, URL count (Walt et al., 2018; Singh et al., 2018; Khaled & El-tazi, 2018; Er et al., 2017; Gilani et al., 2017). After the process of feature extraction, feature selection plays an important role as the accuracy of the trained model depends on the independent variables. Hence, using techniques such as decision tree or Ada-boost SVM which can combine various weak classifiers into a strong model which essentially helps for the sampling of features and implementing models on them (Suganya et al., 2021). In the next section, various classification, clustering, and regression techniques for the detection and mitigation of fake profiles are discussed.

3.3. Machine learning algorithms

a) Classification algorithms:

Support vector machine (SVM): The SVM algorithm is used for classification and regression but it is generally used for the classification of large datasets (Kondeti et al., 2021). It has three components, Hyperplane, Margin, and Support vector. Using the hyperplane a dataset can be classified into a binary group with a boundary separating the data points. This type of algorithm works efficiently for large social media datasets (Suganya et al., 2021; Kondeti et al., 2021). The classification accuracy of SVM depends on the features selected. Hence, to build an accurate model, techniques such as Ada-boost can be used where not only the features are trained with various combinations. Instead, various weak algorithms are combined to create a strong model. When performing quantitative analysis, (Suganya et al., 2021) found that SVM when combined with Ada-boost SVM performed even better in terms of AUC score i.e. AUC= 92.84 with SVM and AUC= 97.16 with Ada-boost SVM. A few studies claimed that SVM performed better when compared with some other supervised

learning algorithms like Naïve Bayes, Logistic Regression, and Random Forest (Suganya et al., 2021; Beskow & Carley, 2019).

b) Clustering algorithms:

K-means and K-medoids: Social media platforms are essentially a structure of nodes which has interconnecting links and associations. The best way to classify a large group of data nodes is to form clusters of the nodes based on similarities with the use of algorithms such as K-Means and K-Medoids (Wang et al., 2019) While selecting the features for modeling, rather than focusing on individual accounts, selection features which could define the whole cluster could create a more accurate model (Xiao et al., 2015). One of the issues of dealing with a large dataset is the volume of the dataset. Trying to label individual profiles is more resource-consuming than labeling clusters of similar profiles as the number of clusters would be lesser than the volume of individual profiles (Wang et al., 2019). A benefit of clustering is an easier representation of large volume data.

c) Regression algorithm:

Logistic regression and Linear regression: There are two types of regression, linear and logistic regression. Linear regression divides the dataset using a straight line into a binary classification (Kondeti et al., 2021). Unlike linear regression, logistic regression uses probability to classify the data points (Belokurov et al., 2021). These techniques can be used to perform classification on large social media datasets. Logistic regression can be used to predict if the social media profile is a bot or not. A study has shown that Logistic regression performed better than Naive Bayes and SVM with an AUC of 0.996. The algorithm performed well even in other factors such as F1 score and precision (Beskow & Carley, 2019).

d) Neural network:

Datasets in social media have a heterogeneous mixture of data with different datatypes. To process data like this, a more powerful approach needs to be introduced. Neural network is a method that can speed up the process of finding insights from a large dataset (Khaled & El-tazi, 2018). Given the fact that data on social media has heterogeneous data types such as videos, photos, texts, and audio. A Convolutional neural network can be used to perform deep learning on such data (Belokurov et al., 2021; Borkar & Sharma, 2020). One of the major advantages of using neural networks is the ability to use hybrid algorithms to make the model more accurate. For example, a neural network processes the decision values from SVM as a hybrid classifier. This method was observed to be efficient as 98% of the dataset was correctly classified from the training dataset (Khaled & El-tazi, 2018). Neural networks when combined with SVM can produce good accuracy. For multiple feature sets like PCA, Correlation, and Regression. The SVM-Neural network gave an accuracy of 0.922, 0.983, and 0.96 respectively (Khaled & El-tazi, 2018).

3.4. Evaluation process

a) Performance metrics and Mathematical models:

Building a machine learning model is not enough on its own, a performance matrix summarises the results obtained by different machine learning models and represents the data in the form of different parameters such as AUC, Recall, precision, and F1 score (Van Der Walt & Eloff, 2018; Walt et al., 2018; Wang et al., 2019; Mohammadrezaei et al., 2018). AUC is a parameter that defines if the classification model has performed well. If the value of AUC is closer to one, it implies that the model performed well (Mohammadrezaei et al., 2018; Kondeti et al., 2021). Along with these parameters TP, FP, FN, TN are also used to check if the classified social media accounts were correctly classified into their expected classes. Based on these values Accuracy is calculated as follows. Accuracy=(TP+TN)/ (TP+FP+TN+FN) (Mohammadrezaei et al., 2018; Kondeti et al., 2021).

b) Data visualization:

The final step of building a machine learning model is to visualize the data points classified by the machine learning models. For that, various types of data visualization models can be used but it depends on the type and volume of the data which needs to be visualized. A large volume of datasets on social media with diverse datatypes is a major concern while representing data graphically (Khaled & El-tazi, 2018). To represent a large amount of data effectively, reducing dimensions of the data can reduce the irrelevant features hence making it easier to visualize (Khaled & El-tazi, 2018). To visualize classification data produced by the models Confusion matrices and learning curves can be used for efficient visualization based on the frequency of results (Narayanan, 2021).

4. Gaps and limitations

4.1 Observations:

After studying numerous research content, it seems like there are some obvious gaps between the existing Fake ID detection techniques. Previously done research lacks in the scope of using the ML techniques on all platforms. It looks like there are no standardized solutions for detecting fake profiles universally for social media platforms. The problem of ID deception is spread across every social media platform which needs to be addressed to reduce ID frauds and other cyber security threats. While working on big datasets, another challenge that needs to be faced is dealing with a large number of data points. Even though a solution to this problem was addressed using the clustering technique (Xiao et al., 2015) the limitation of this technique is that it can only find a platform-oriented solution cause the features might differ from platform-to-platform bases. There are different types of actors like humans and bots which can carry out ID fraud. To distinguish between them a solution would be to use features that can clearly differentiate bots from human actors. It can be done with the help of engineered

features (Van Der Walt & Eloff, 2018) however, there is a limitation to this technique. Not all platforms can use the same engineered features to overcome this problem as the data and the data storing methods changes from site to site. In the case of human actors, the process of falsifying identity is related to human psychology (Walt et al., 2018). This problem can be addressed by choosing appropriate features which can demonstrate human traits. However, this technique has a limitation that the selected features can raise an ethics concern of selection bias which can lead to an increase in FP rates.

5. Conclusion

After analysing different techniques of machine learning for detecting fake identities on social media, there are a few machine learning models which performed better than the others. A lot of factors contribute to building an efficient model. Choosing relevant features which can be used on various platforms to build a universal solution would be an economic choice. While selecting the features, human psychology should be considered to distinguish between bot actors and human actors. Social media datasets can have a lot of dimensions that are often hard to work with. A solution for this problem is to use Principle component analysis (PCA) to reduce the dimensionality and feed those features to a machine learning model. If these features are used for Support vector machine classifier when combined with Ada-boost can yield promising results. To further improve the classification process Neural Network (NN) can be used to compute different algorithms with less computing resources and greater efficiency. However, all of these techniques when implemented to a cluster of social media accounts that are grouped by similarities, instead of individual accounts can increase the efficiency and computing capacity as social media users are increasing exponentially and so are fake profiles with the corresponding cyber security threats.

6. References

Belokurov, D. A., Shamakova, E. S., & Kolomoitcev, V. S. (2021). *Using Machine Learning Techniques to Identify Bot Accounts on a Social Network*. 1–5. https://doi.org/10.1109/weconf51603.2021.9470605

Beskow, D. M., & Carley, K. M. (2019). Its all in a name: detecting and labeling bots by their name. *Computational and Mathematical Organization Theory*, *25*(1), 24–35. https://doi.org/10.1007/s10588-018-09290-1

Borkar, B. S., & Sharma, M. (2020). *Identification of Fake Identities on Social Media using various Machine Learning Algorithm*. *9*(4). https://doi.org/10.30534/ijatcse/2020/299942020

Er, B., Akta, Ö., Deniz, K. Õ. Õ., & Akyol, C. (2017). *Twitter Fake Account Detection*. 2–6. https://doi.org/10.1109/UBMK.2017.8093420

Gilani, Z., Kochmar, E., & Crowcroft, J. (2017). Classification of twitter accounts into automated agents and human users. *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2017*, 489–496. https://doi.org/10.1145/3110025.3110091

Khaled, S., & El-tazi, N. (2018). *Detecting Fake Accounts on Social Media*. *October 2017*, 3672–3681. https://doi.org/10.1109/BigData.2018.8621913

Kondeti, P., Yerramreddy, L. P., Pradhan, A., & Swain, G. (n.d.) (2021). *Learning*. Springer Singapore. https://doi.org/10.1007/978-981-15-5258-8_73

Mohammadrezaei, M., Shiri, M. E., & Rahmani, A. M. (2018). Identifying Fake Accounts on Social Networks Based on Graph Analysis and Classification Algorithms. *Security and Communication Networks*, *2018*. https://doi.org/10.1155/2018/5923156

Narayanan, A. (n.d.) (2018). *IronSense : Towards the Identification of Fake User-Profiles on Twitter Using Machine Learning Department of Computer Science*. https://doi.org/10.1109/ICINPRO43533.2018.9096687

Singh, N., Sharma, T., Thakral, A., & Choudhury, T. (2018). *Detection of Fake Profile in Online Social Networks Using Machine Learning*. 231–234. https://doi.org/10.1109/ICACCE.2018.8441713

Suganya, R., Muthulakshmi, S., Venmuhilan, B., Varun Kumar, K., & Vignesh, G. (2021). Detect fake identities using improved Machine Learning Algorithm. *Journal of Physics: Conference Series*, *1916*(1), 012056. https://doi.org/10.1088/1742-6596/1916/1/012056

Van Der Walt, E., & Eloff, J. (2018). Using Machine Learning to Detect Fake Identities: Bots vs Humans. *IEEE Access*, *6*, 6540–6549. https://doi.org/10.1109/ACCESS.2018.2796018

Walt, E. Van Der, Eloff, J. H. P., & Grobler, J. (2018). Cyber-security : Identity deception detection on social media platforms. *Computers & Security*, *78*, 76–89. https://doi.org/10.1016/j.cose.2018.05.015

Wang, X., Lai, C. M., Lin, Y. C., Hsieh, C. J., Wu, S. F., & Cam, H. (2019). Multiple Accounts Detection on Facebook Using Semi-Supervised Learning on Graphs. *Proceedings - IEEE Military Communications Conference MILCOM*, *2019-Octob*, 94–101. https://doi.org/10.1109/MILCOM.2018.8599718

Xiao, C., Freeman, D. M., & Hwa, T. (2015). Detecting clusters of fake accounts in online social networks. *AISec 2015 - Proceedings of the 8th ACM Workshop on Artificial Intelligence and Security, Co-Located with CCS 2015*, 91–102. https://doi.org/10.1145/2808769.2808779