ZULFAQAR Journal of
Defence Science, Engineering & Technology

Journal homepage: https://zulfaqar.upnm.edu.my/

# CLASSIFYING FAKE PROFILE IN FACEBOOK ACCOUNT USING SUPPORT VECTOR MACHINE

**Ahmad Nazren Hakimi Ahmad Nasir, Suzaimah Ramli*, Muslihah Wook, Noor Afiza Mat Razali, Norulzahrah Mohd Zainuddin**

*Department of Computer Science, Faculty of Science & Defence Technology, National Defence University of Malaysia, Sg. Besi Camp, Kuala Lumpur, Malaysia*

*Corresponding author: *suzaimah@upnm.edu.my*

**ARTICLE INFO**

***Article history:***
Received
*09-07-2020*
Received in revised
*30-10-2020*
Accepted
*06-05-2021*
Available online
*31-12-2021*

***Keywords:***
Facebook, Fake profile, Machine learning, Online social media, SVM

e-ISSN: 2773-5281
Type: Article

**ABSTRACT**

*The ability to connect people around the world and share videos, photos and communications has made online social media (OSM) popular. Despite much of literature available in this field, there is still lack of study focusing on the creation of fake profile in the OSM. In fact, there has been relatively little effort aimed at solving fake profile features using classification algorithm. For these reasons, Support Vector Machine (SVM) classifier was employed to classify the novel features that had been created in the fake profile particularly among Facebook users' account. The study begins with the data collection process, data pre-processing, evaluation, testing and lastly obtaining the result. The findings have revealed that the SVM classifier able to predict the fake profile with high Classifying Accuracy (CA) and Area Under Curve (AUC). Ultimately, this finding will provide a new endeavour for countermeasure and protection of OSM users.*



## Introduction

Digital technology has revolutionized the Internet and social networks, influencing everyday life as well. A person has a simulated world very closely linked to everyday life. Friendships are often developed via Facebook, work can be found via LinkedIn, and the Internet provides more social networking possibilities. The world has changed, and technology allowed us to enter a virtual universe where many critical things can be found and accomplished. Users recently displayed massive engagement on social networking platforms, culminating in a steady speed of user-generated content. For, e.g., Google searches over 100 billion a month, and Google indexes more than 50 billion websites. Meanwhile, Twitch has over 1 billion Internet users, a quarter of all internet consumption. These people upload videos per minute for 100 hours on average [1].

Facebook also is a big online social network (OSN). This is used to share information with other users and celebrities, politicians, and other persons of public concern. The site has 1.3 billion users, spending 640 million minutes on 54 million sites a month [2]. In social media, users' power and popularity play a central role; many people imitate and trust prominent accounts. What happens if the trusted account is fake. Fake profile are social network accounts created and provided to users to maximize their popularity and participation for different social motives. Fake profile, however, are only one example of "anomalous" identities spreading through social networks [3-4].

In this article, one classification algorithm was used by running the Support Vector Machine (SVM) on a collection of real Facebook social media data to differentiate between a fake and a real Facebook account. Besides, we boost the grouping parameter by integrating online signatures, such as the Internet Protocol (IP) and the login page. The remainder of this paper is structured as follows. The literature review offers an overview of work undertaken on Facebook and previous studies on fake profile identification. In terms of methodology, Facebook datasets have been identified that illustrate how the data gathered has been reprocessed and used to identify accounts in bogus accounts and actual accounts. As a result, the average performance ratings have been explored and contrasted with all other approaches used. Lastly, we are presenting our discussion and conclusion based on some perspectives.

## Literature Review

*Online Social Media (OSM)*

OSM is one of a place that people can have interaction without going out from their home bases. They can share about their activities to other, pictures, messages and video all over the world. Based from Vošner et al. (2016), it is possible to use online social networks to communicate with people regardless of time or location [5]. Social media generally relies on computer-mediated communication and is characterized as tools and platforms for the consumption, co-creation, sharing and modification of user-generated content. Social media applications can be used to interact with other people via blogs, content communities, social networking sites, virtual game worlds or social worlds.

Sharing our particular privacy information absolutely or partially can exposed to the public makes us ideal candidates for specific attack forms, the worse of which may be identity fraud. Identity theft happens when some person uses the skill of character for a private reason. Online identity fraud has been a critical concern in the earlier years, as it impacted millions of people worldwide. Victims of identity fraud may suffer unusual forms of penalties, they may forfeit time or cash, for example, be sent to a jail, destroy their public reputation, or harm their relationships with friends and loved ones [6]. Offers in social networking also encouraged identity stealing and assaults on impersonation by both extreme and innocent perpetrators [7].

*Fake Profile Characteristics*

By sharing our identity in social network, it may be vulnerable to identity fraud. Most of the scammer like to steal personal identity of real account in social network and use it for bad things. Here we will discuss about the fake profile characteristic that had been used to differentiate between fake profile and real profile. According to study in [4, 9], they had listed the features characteristic of fake profile that been used and the list in Table 1:

**Table 1: List of Fake Profile Characteristics**

| | **Author** | |
|---|---|---|
| **Characteristics** | **(Kudugunta & Ferrara, 2018)** | **(Viswanath et al. 2014)** |
| Status count | √ | √ |
| Friend count | √ | √ |
| Follower count | √ | x |
| Favourites count | √ | x |
| List count | √ | x |
| Default profile | √ | x |
| Likes count | √ | √ |

| | Author | |
|---|---|---|
| **Characteristics** | **(Kudugunta & Ferrara, 2018)** | **(Viswanath et al. 2014)** |
| Comment count | √ | √ |
| User activities | √ | √ |
| Tags count | x | √ |
| Share count | x | √ |
| Geo enable | √ | √ |
| Profile uses Background image | √ | x |
| Protected | √ | x |
| Verified | √ | x |

When following these lists above, we need to minimize the selection of the features characteristic above. There are two main reasons for the option of minimization the size of the feature set. First, Model efficiency: A reduced range of features gives very efficient models that can be trained more easily and are less likely to over fit, a common problem in social media data mining due to the existence of outliers. Second, Interpretability: A limited set of features with obvious significance such as those provided by account metadata enables interpretable models to be produced. This is important matter, especially when combined with notoriously difficult to interpret deep learning strategies [9].

*Focusing on Facebook*

Online Social Networks (OSN) have also attracted researchers for mining and analysing their vast volume of data, investigating and researching consumer habits, and identifying suspicious activities [10]. Besides researching fake accounts from a technical perspective, studies of behavioural patterns are often required. OSN user activity includes numerous online networking behaviours such as friend formation, content posting, profile searching, chatting, and commenting. This information can be divided into two categories, public information and private information [11]. Researchers used web crawlers mainly to collect public knowledge from OSNs users. Such web crawlers can acquire user information by using the OSN programming interface (API) or by analysing raw data gathered directly from the OSN web pages. However, some OSNs like Facebook cannot gather public information from users without being signed into the OSN. To overcome this restriction, researchers have built passive fake profiles used to obtain access to public OSN information [12].

These false profiles don't trigger friend requests to other network users and don't interfere with OSN activity. Researchers developed many methods to collect private information from OSN users. These tactics involve demanding private information directly from users via applications and software add-ons that interact with the OSN [13]. Inferring private information from OSN users through examining details gathered from their contacts and also triggering dynamic fake accounts, also known as social bots, that trigger a set of friend requests to cooperate [14]. Using these techniques, researchers may gain a more detailed description of the analysed OSN, including knowledge about private users.

Based on the features that been discuss in section 2.2, we are focusing on classifying Facebook fake profile. This is because Facebook dataset are compatible with features listed above. Other than that, according to Institut Penyelidikan Pembangunan Belia Malaysia, MCCA and Multimedia, up to 97.3% of Malaysian citizen are using Facebook on 2019 [15-17]. By this, tendency of creating fake profile on Facebook are high in Malaysia. That why we need to focusing on Facebook fake profile because some of them create fake profile and use it as a bot for some reason. Some are benevolent and, in theory, harmless or even helpful: this group involves bots that automatically collect information from multiple sites, including basic news feeds [18]. Brands and businesses gradually implement automated inquiries responders for customer service. While such forms of bots are intended to have a helpful service, they may also be dangerous, for example, if they help propagate unverified knowledge or rumours.

*SVM Classification*

There are two primary methods known as Supervised Learning in machine learning. The training dataset has a class name, and Unsupervised Learning, where data are clustered together based on measurable actions or characteristics. In other words, a defined collection of training data is used to approximate or map the input data to the target output. In comparison, no labelled instances are given under unsupervised

methods. There is no notion of performance during the operation. Instead, data with similar attributes or similar actions are grouped (clustered) [19]. To detect these fake profiles by using features that been identify, we propose to concentrate on using the Support Vector Machine (SVM) technique. SVM is a decision plane idea that fines the limit of a decision.

The SVM aims to locate a hyperplane in the sum of characteristics that precisely identify the data point. It is mainly a classifier technique that executes functions in a multidimensional space by constructing hyperplane that separates instances of various class marks. SVM may have multiple constant and categorical variables. SVM promotes regression and grouping [20]. SVM also applies the concept of structure risk minimization that minimizes both analytical error and learner uncertainty and achieves success in classification and regression tasks. SVM's classification aim is to create the optimum hyperplane with a full margin. The larger the gap, the lower the classifier's generalization error. Data are classify using SVM will be evaluate based on accuracy, Area Under Curve (AUC) and ROC.

SVM has two forms of designation, C-SVM and V-SVM. We need to pick what sort of SVM needs to be used. It is based on the form of classification for the test error settings. C-SVM and v-SVM are based on specific error minimizations. We will set the check error limits, the expense for C-SVM and the complexity limit for v-SVM. Identify SVM kernel also play important role to obtain good result. The next block of options deals with the kernel, a function that converts the attribute space to a new feature space to suit the maximum-margin hyperplane, allowing the algorithm to construct non-linear classifiers for the Polynomial, RBF and Sigmoid kernels. The functions that define the kernel are described in addition to the names of the kernel, and the constants concerned are:

g - for the gamma constant in the kernel function (the suggested value is 1 / k, where k is the number of attributes, but because there can be no training set for the widget, the default value is 0, and the user has to configure this option manually).
c - for the constant c0 of the kernel function (default 0) and
d - for the kernel stage (default 3).

Some of scholar had proven using SVM classifier predict fake profile with high accuracy, Area Under Curve (AUC) and low false positive rate [18-21].

## Methodology

Based on the feature's characteristics of a fake profile detection, this segment presents our proposed method. It starts from data collection, then pre-processing process, evaluation and testing dataset and lastly result obtain.

*Data Collection*

This research demands real-world Facebook databases not publicly accessible. There are several social network databases available with profile-based feature data, but these databases are anonymized and impossible to use. Therefore, the analysis must collect data from the Facebook API, as it is restricted to registered users. When Facebook is continually changing privacy policies, it is also impossible to access data without Facebook permission. This study used generated data from an online platform [20]. To overcome this problem, research has been conducted by focusing on a dataset by the student of Universiti Pertahanan Nasional Malaysia (UPNM). We randomly collect 100 student's Facebook datasets with their permission and concern as a sample of real-world Facebook datasets. All the datasets are saved as JSON format. JSON is a compact, hierarchical format for parsing and processing in many programming languages [22]. The process flow on research methodology, as shown in Fig. 1.

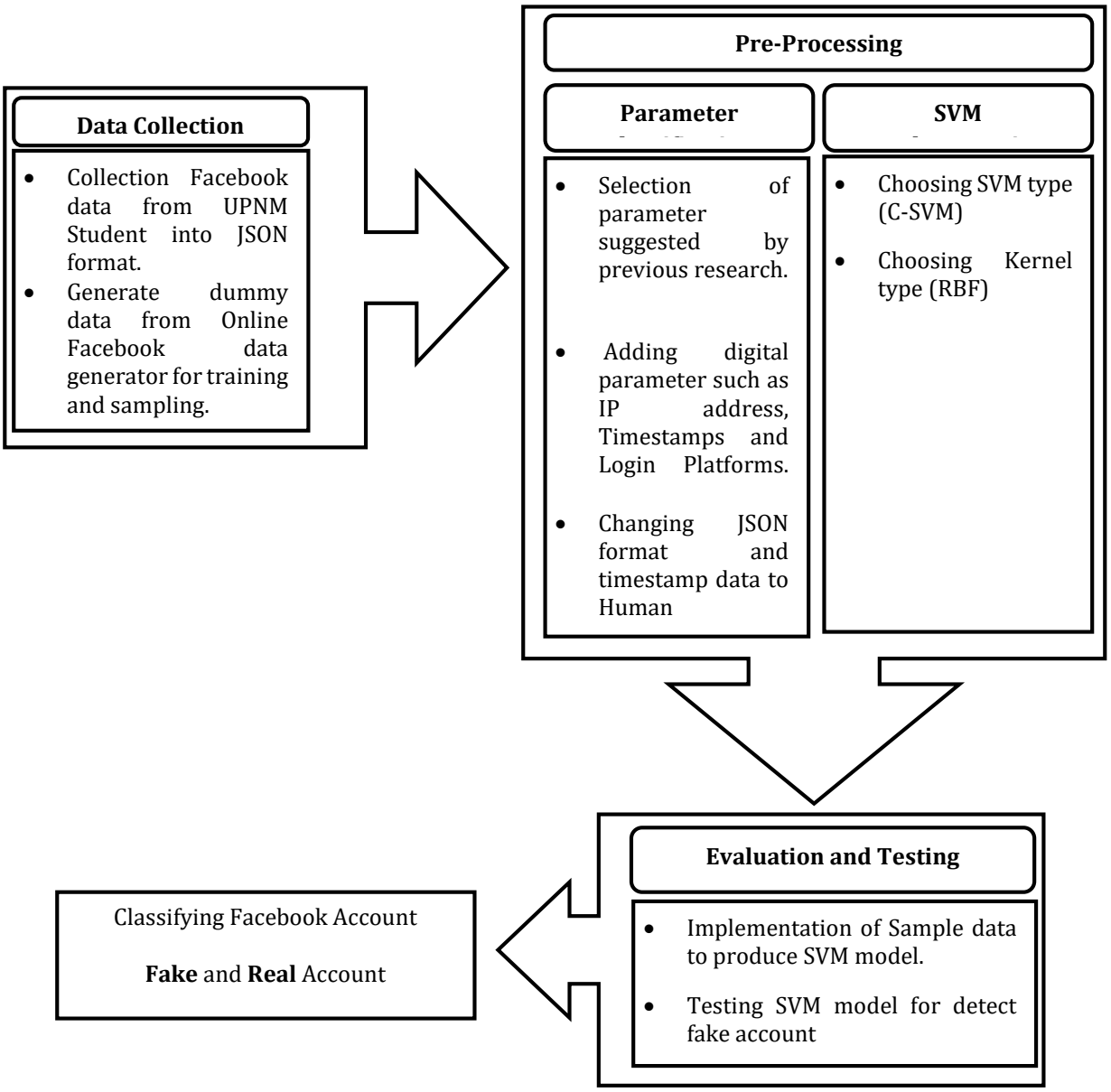


**Fig. 1: Implementation of process flow**

*Parameter identification*

Based on research done, most of the research only focusing on the pattern of behavior of fake profiles. Section 2.2 had discussed about fake profile characteristics and combine these characteristics with our previous study had simplified the features into five features [20]. In this analysis, we propose to use some of the characteristics and added new characteristics into our investigation by combination of the digital signature and the sequence behavior of the fake profile identified in the datasets. Apart from five features from the previous study, we added another two digital features. First, we use timestamps to determine the pattern of activity.

Timestamp or timestamp is the time stored in a file, log, or message that tracks when data is inserted, removed, changed, or transmitted and used Unix and Epoch format [22]. We need to convert it to the human-readable data format [23]. It would make the result firmer and more credible. Second, we used their login platform and IP address to evaluate how they could connect to their fake profile. From this, we can differ either user of a fake profile using gadgets such as smartphones or computers based on their artifact data. It will help the forensic matter later [24] After reviewing all the features and parameter, Table 2 is the list that will be used in this study.

**Table 2: Parameter use**

| Features selection | Justification |
|---|---|
| Friend Count | Real profile may have many real friends and have bigger interaction activities rather than fake profile. |
| Status/ Post/Comment activity Count | Fake profile expected to post and share spam content and get small amount of comment. |
| List/Like Count | Fake profile might be more active than real account. |
| Timestamp (Proposed parameter) | Every detail of user activities is recorded to analyses their pattern. Fake profile expected to has more suspicious activity. |
| Usage of IP and login platform (Proposed parameter) | Fake profile expected to has more suspicious IP address and login platform due to usage of proxy tool. |

*Implementation of SVM*

We used C-SVM for non-complex classification and Radial Basis Function kernel (RBF) in this research because it is suitable for a medium dataset. All the settings above are shown in Fig. 2

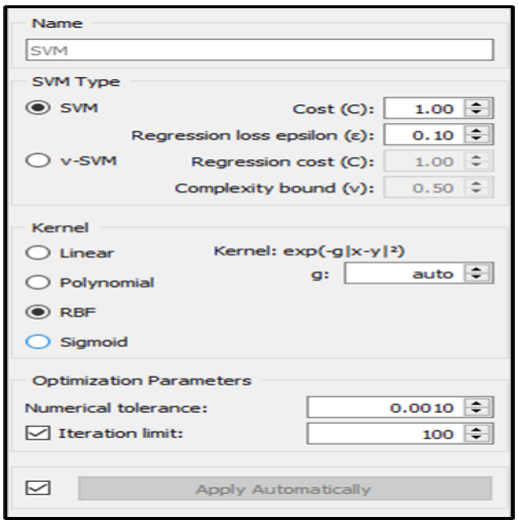


**Fig. 2: Setting of SVM implementation**

We must split all datasets into a training set and test set before you can provide the SVM with the data that has been loaded for predictive analytics. By this, we use generated data from [20]; and make some changes to suit the real-world data. After a model of predictive analysis has been developed, real Facebook datasets can be evaluated. The findings were compiled and compared in another segment below. Past research had shown that SVM offers reasonable reading on Clarity Accuracy (CA) and Area Under Curve (AUC), which we can use to justify fake profiles.

## Results & Discussion

This section provides an analysis of the techniques used to determine the output of fake profiles detected on Facebook. A combination of data sets consisting of previously recognized real profile from the first and second phases of the users of the social district and the Fake profiles were introduced to all of the above classificatory. The dataset also contains user accounts friends of social neighborhood colleagues, supposed to be actual in active accounts and supposed to be fake inactive. The evaluation is based on confusion matrix and associate matrix [18]. The variable of True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN) in the confusion matrix at Table 2 is refer to following:

True Positive (TP): number of fake profiles that are identified as fake profile. (Fake + Fake)
False Positive (FP): number of real profiles that are identified as fake profile. (Real + Fake)
True Negative (TN): number of real profiles that are identified as real profile. (Real + Real)
False Negative (FN): number of fake profiles that are identified as real profile. (Fake + Real)

During determining the true positive rate (TPR) and false positive rate (FPR), we used several formulas to calculate.

$$\text{True Positive rate} = \frac{\text{No of fake profile detected}}{\text{Total number of fake profiles}}$$

$$\text{False Positive rate} = \frac{\text{No of real profile detected as real profile}}{\text{Total number of real profiles}}$$

Since the application of the SVM Algorithm, we find that the Accuracy Classification is 0.807. It is better than those in previous research [20]. So, the accuracy of the prediction is almost 80 percent. Any of the data were missing or had any irregularities during the classification process. These are overcome by imputing all incomplete information to random values. This type of data on defects is needed in the research to prevent bias from occurring. Table 3 shows the output of the SVM classifier.

**Table 3: Result of SVM Classifier**

| Model | AUC | CA | F1 | Precision | Recall |
|---|---|---|---|---|---|
| SVM | 0.95 | 0.85 | 0.85 | 0.85 | 0.85 |

Based on Fig. 3, the SVM algorithm capable of identifying a fake profile based on the parameter was present and seen in the scatter plot graph. Fig. 3 is an example of the results of the research.

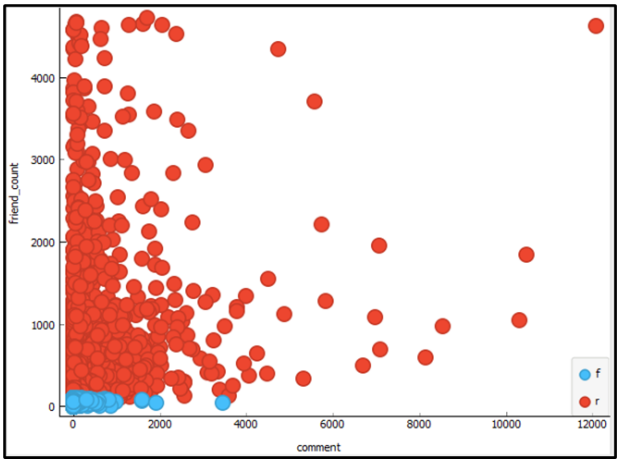


**Fig. 3: Example of result data**

The data is pretty good for the new parameter that we introduced. It can be distinguished from a fake profile of above 80%. The table below provides a set of estimates based on the data we offer. Result in Table 3 are illustrated as in Fig. 4 below.

**Table 3: Result of Predicted**

| | | **Predicted** | |
|---|---|---|---|
| | | **Fake** | **Real** |
| Actual | Fake | 86.3% | 15.9% |
| | Real | 13.7% | 84.1% |

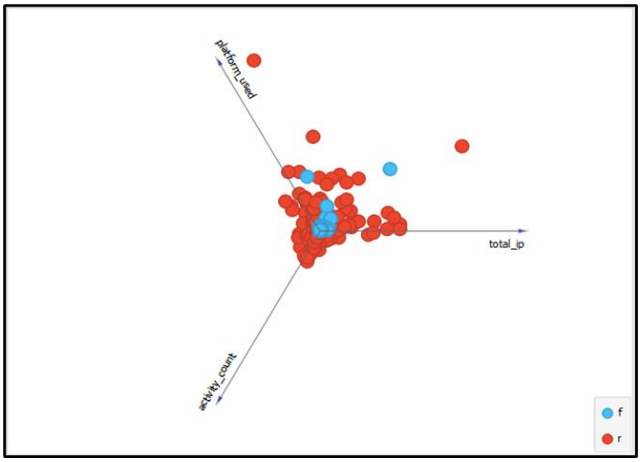


**Fig. 4: Detection on new parameter added**

## Conclusion

Through the years, fake profiles have continuously developed to keep them from being found. It is, therefore, important to establish methods for identifying fake profiles. This analysis shows the basics of the hunt for false Facebook profiles focused on users from university students based on user profile activities and contact with other users on Facebook. The study used artificially generated and real university student datasets for Facebook features, as the fine-grained privacy settings on Facebook posed a major challenge to data collection. The most commonly used forms of machine learning classification are then used to classify the fake profile. After review show that SVM algorithm can detect fake profiles of up to 80% accuracy. However, the fake profile happens in student use as a social platform rather than do bad things but still have opportunities to engage in cybercrime activities. If a fake profile is real and happens, studying how to overcome it and its countermeasure to detain this account need to be emphasized in future work for strengthen the security of online social media user's.

## References

[1] Ramalingam, D., & Chinnaiah, V., "Fake profile detection techniques in large-scale online social networks: A comprehensive review," *Computers and Electrical Engineering*, Vol. 65, No. 3, 2018, pp. 165–177.

[2] Fire, M., Çalişmalari, Y. Z., Şİmşek, M., Yilmaz, O., Kahrİman, A. H., Sabah, L., Gheewala, S., Patel, R., Siewert, B. S., Siewert, S., M. Meligy, A., M. Ibrahim, H., F. Torky, M., Romanov, A., Semenov, A., Veijalinen, J., Al-Qurishi, M., Al-Rakhami, M., Alamri, A., Weippl, E., "A sneak into the Devil's Colony - Fake Profiles in Online Social Networks," *Journal of Supercomputing*, Vol. 5, No. 1, 2018, pp. 26–39.

[3] Caruccio, L., Desiato, D., & Polese, G., Fake Account Identification in Social Networks. *Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018*, 2018, pp. 5078–5085.

[4] Viswanath, B., Bashir, M. A., Crovella, M., Guha, S., Gummadi, K. P., Krishnamurthy, B., & Mislove, A., Towards Detecting Anomalous User Behavior in Online Social Networks. *Proceedings of the 23rd USENIX Security Symposium*, 2014, pp. 223–238.

[5] Vošner, H. B., Bobek, S., Kokol, P., & Krečič, M. J. (2016). Attitudes of active older Internet users towards online social networking. *Computers in Human Behavior*, Vol. 55, 2016, pp. 230–241.

[6] Rao, P. S., Gyani, J., & Narsimha, G., "Fake Profiles Identification in Online Social Networks Using Machine Learning and NLP," *International Journal of Applied Engineering Research ISSN*, Vol. 13, No. 6, 2018, pp. 973–4562.

[7] Kamoru, B. A., Bin, A., Omar, J., & Jabar, M. A., *Spam Detection Issues and Spam Identification*. Vol. 95, No. 21, 2017, pp. 5881–5895.

[8] Kudugunta, S., & Ferrara, E., "Deep neural networks for bot detection," *Information Sciences*, Vol. 467, pp. 312–322.

[9] Krombholz, K., Merkl, D., & Weippl, E., "Fake identities in social media: A case study on the sustainability of the Facebook business model," *Journal of Service Science Research*, Vol. 4, No. 2, 2013, pp. 175–212.

[10] Kaur, R., & Singh, S., "A survey of data mining and social network analysis based anomaly detection techniques," *Egyptian Informatics Journal*, Vol. 17, No. 2, 2016, pp. 199–216.

[11] Elovici, Y., Fire, M., Herzberg, A., & Shulman, H., "Ethical Considerations when Employing Fake Identities in Online Social Networks for Research," *Science and Engineering Ethics*, Vol. 20, No. 4, 2014, pp. 1027–1043.
[12] Fire, M., & Puzis, R., "Organization Mining Using Online Social Networks," *Networks and Spatial Economics*, Vol. 16, No. 2, 2016, pp. 545–578.
[13] Elishar, A., Fire, M., Kagan, D., & Elovici, Y., Organizational intrusion: Organization mining using socialbots, *Proceedings of the 2012 ASE International Conference on Social Informatics, SocialInformatics 2012*, *SocialInformatics*, 2012, pp. 7–12.
[14] Fire, M., Kagan, D., Elishar, A., & Elovici, Y., *Sotics_2012_3_10_30021.Pdf. c*, 2012, pp. 46–50.
[15] Institut Penyelidikan Pembangunan Belia Malaysia, *Institut Penyelidikan Pembangunan Belia Malaysia,* (2020). https://www.iyres.gov.my/index.php/statistik-pengguna-internet-facebook
[16] MCCA, *Persatuan Pengguna Siber Malaysia,* 2020. https://www.cyberconsumer.my/penerbitan/statistik-pengguna-internet-di-malaysia-2018/
[17] Multimedia, K. D. A. N., *Cabaran komunikasi dan multimedia*. Vol. 41, 2019.
[18] Mohammadrezaei, M., Shiri, M. E., & Rahmani, A. M., "Identifying Fake Accounts on Social Networks Based on Graph Analysis and Classification Algorithms," *Security and Communication Networks*, 2018, pp. 1–8.
[19] Albayati, M. B., & Altamimi, A. M., "An empirical study for detecting fake facebook profiles using supervised mining techniques," *Informatica (Slovenia)*, Vol. 43, No. 1, 2019, pp. 77–86.
[20] Hakimi, A. N., Ramli, S., Wook, M., Mohd Zainudin, N., Hasbullah, N. A., Abdul Wahab, N., & Mat Razali, N. A., "Identifying Fake Account in Facebook Using Machine Learning," In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics),* Vol. 9429, 2019, pp. 441–450.
[21] Syed Mansoor, S. B. R., Halip, M. H. M., Azahari, M. A., Kamarudin, N. D., Mohamed, H., "New Traceability Approach Using Swarm Intelligence For Military Blockchain," *Zulfaqar Journal Of Defence Science, Engineering & Technology*, 2021.
[22] Computer Hope, *Timestamp*, 2019. https://www.computerhope.com/jargon/t/timestam.htm
[23] Misja.com., *Epoch & Unix Timestamp Conversion Tools,* 2020. https://www.epochconverter.com/
[24] Norouzizadeh, D. F., Dehghantanha, A., Eterovic-Soric, B., & Choo, K. K. R., "Investigating Social Networking applications on smartphones detecting Facebook, Twitter, LinkedIn and Google+ artefacts on Android and iOS platforms," *Australian Journal of Forensic Sciences*, Vol. 48, No. 4, 2016, pp. 469–488.