Fake account detection in social media using machine learning methods: literature review

Nalia Graciella Kerrysa[1], Ika Qutsiati Utami[1,2]
[1]Data Science Technology, Department of Engineering, Faculty of Advanced Technology and Multidiscipline, Universitas Airlangga, Surabaya, Indonesia
[2]Graduate Institute of Network Learning Technology, National Central University of Taiwan, Taoyuan, Taiwan

Article Info

Article history:

Received Nov 26, 2022
Revised Apr 5, 2023
Accepted Jun 4, 2023

Keywords:

Fake account
Machine learning
Social media
Society
Technology

ABSTRACT

With the rapid development of emerging technologies in the industrial revolution 4.0 or 5.0, social media has become one of the social environments to carry out social activities, both socializing and advertising. However, since it is an open platform by nature, cybercrime occurrence in social media is inevitable. Currently, more than a million fake accounts are existing on Instagram, Twitter, and Facebook, intending to increase followers, spread hoaxes, and spam. On one hand, it is difficult to manually eliminate these accounts on social media platforms. On the other hand, research on automatic fake account detection has been carried out for more than a decade. This study provides literature reviews aiming to deliver information about several methods and machine learning algorithms with the performances measured in identifying fake accounts on three well-known social media platforms: Twitter, Instagram, and Facebook.

Corresponding Author:

Ika Qutsiati Utami
Data Science Technology, Department of Engineering
Faculty of Advanced Technology and Multidiscipline, Universitas Airlangga
Dr. Ir. Soekarno Rd., Surabaya, Indonesia
Email: ika.q.utami@ftmm.unair.ac.id

1. INTRODUCTION

In the era of industry 4.0, social media has become one of the most preferable platforms to socialize and connect with other people. Interactive social media platforms such as Instagram, Twitter, and Facebook have become popular over a decade to share and find important information [1], [2]. People can connect with others around the world without limitations of time and place. Currently, the purpose of social media interaction is not just for communication and online interaction, but also for conducting business activities such as advertising, promoting, and doing campaigns. Meanwhile, the government could use these platforms to deliver government services to citizens effectively [3]. All of these activities require a lot of followers to be engaged with meaningful interaction to achieve a profitable business purpose. To achieve this, it is inevitable for business people or a company to utilize fake accounts on social media intentionally. Fake accounts are used in many different ways. Most businesses and institutions today choose social media as their main platform for marketing and advertising campaigns [4]. Meanwhile, influencers receive many tangible profits from endorsing brands and sponsorship [5], [6]. Both cases need a huge amount of followers and finding fake accounts provides the fastest solution for a bigger profit. Although they seemed to have no significant impact, fake accounts could also run a lot of devious activities over the internet such as launching massive online attacks [7], spreading hoaxes, review-bombing of products with misleading content, spreading spam, and even impersonating someone [8]. Not only that cases, these fake accounts could take advantage of people by faking news or text messages to steal from innocent social media users [9], [10]. These activities will affect the reputation of individuals or groups of people on a larger scale [11]. One example is when a 17-year-old high school student's identity was stolen by a large company that sells Twitter followers to anyone that wants to become popular [12]. The company claimed approximately 3.5 million automated accounts including stolen identities from which they profited. Another case is when Twitter faced a problem where fake accounts would get verified which slows down the verification of important and official accounts since the process needs to be improved [13]. As mentioned by Elyusufi *et al.* [11], the existence of fake accounts is considered more dangerous than any other cybercrimes. Moreover, the existence of fake accounts in social media can't be tracked and removed easily. We need some techniques to automatically solve this problem. On the other hand, with the advancement of technology and algorithms, some fake accounts are trained to mimic the activities of a real social media user so that they can avoid deletion from the respected social media platform [14]. People also make some traditions to purchase many fake accounts with affordable prices to meet their business purposes [15]. This condition will lead to an increasing number of fake accounts over time. The inability to minimize these fake accounts automatically and effectively is the main reason for current research topics focused on this part. This literature review aims to summarize some of the research studies focusing on machine learning techniques to detect social media fake accounts. This study also provides information for a high-performance model that can be implemented in detecting fake accounts on Instagram, Facebook, and Twitter.

2. METHOD

The method used for the literature review of this study refers to the method conducted by Zuhroh and Rakhmawati [16] which consists of 4 steps that include defining research questions, literature keywords and sources, study criteria selection, and the findings of the literature study. We also followed a guide to structure a literature review by Kitchenham and Charters [17]. This literature review included 4 stages as follows:

a. Setting up the literature review goals and questions

The objective of this stage is to find methods for fake account detection with better performance on 3 social media platforms e.g., Instagram, Twitter, and Facebook. We compose the research questions as follows:

- What are the attributes or features that can be used to effectively detect fake accounts in social media?
- What are machine learning methods commonly used in fake account classification tasks?
- What is the performance of each machine learning method in detecting fake accounts in social media?

b. Research article selection

The keywords such as "machine learning", "online social network", "social media", "fake account detection", and "fake account classification" are used to search for some related articles from the database i.e., Google Scholar, IEEE, and Scopus. Articles regarding literature reviews are excluded from this study. The search process obtained 30 articles which are shown in Table 1.

Table 1. The searching results

Publication year	Authors	Total
2015	Cresci *et al.* [18]	1
2017	Ersahin *et al.* [8], Gupta and Kaushal [7], Khalil *et al.* [19]	3
2018	Walt and Eloff [20], Raturi [21], Chen and Wu [22]	3
2019	Elyusufi *et al.* [11], Reddy [23], Akyon and Kalfaoglu [24], Singh and Banerjee [25], Khaled *et al.* [26], Mohammad *et al.* [27], Pakaya *et al.* [28]	7
2020	Jabardi and Hadi [29], Purba *et al.* [15], Sheikhi [1]	7
2021	Meshram *et al.* [14], Heidari *et al.* [30], Wang *et al.* [31], Kesharwani *et al.* [32], Bharti and Pandey [33], Narayan [34], Pashwan and Ravi [35]	3
2022	Chakraborty *et al.* [36], Das *et al.* [37], Kadam and Sharma [38], Shreya *et al.* [39]	4
2023	Durga and Sudhakar [40], Reddy *et al.* [10]	2
Total		30

c. Discussion

From the collected articles, three aspects of the research will be discussed. The first is to discuss the dataset used in the study. The second is to discuss the attributes selected by the study. The final task is to discuss the machine learning model used in the study.

d. Data synthesis

After the discussion, the data from the respected study will be elaborated and summarized. The performance of the models will be mentioned with performance metrics used in the article. Aditionally, the evaluation of the results will be explained.

3. RESULTS AND DISCUSSION

Several studies performed multiple steps to obtain the best model for fake accounts classification including: i) gathering datasets in several ways e.g., manually, automatically, and using an existing dataset; ii) applying feature selection for increasing the effectiveness and efficiency of the model; iii) selecting a machine learning model to classify fake accounts; and iv) measure the performance of the model as well as evaluate the result.

3.1. Dataset

Two types of datasets can be used for fake account detection: free datasets and self-made datasets. The majority of researchers chose to make their datasets that comprise fake and real accounts. One of the reasons to build their dataset is because of no available public open datasets for detecting fake accounts [24]. Some researchers collected survey data using a questionnaire [21] or even hired a company to make a part of the dataset [23]. For fake account classification, the dataset is collected from Facebook, Instagram, and Twitter (as shown in Tables 2 and 3). Other platforms are excluded from this literature review.

Table 2. The original dataset

Social media	Reference	Sample size
Facebook	Singh and Banerjee [25]	Fake accounts: 537 Real accounts: 418
	Reddy [23]	1,162 accounts
	Gupta and Kaushal [7]	4,708 accounts
	Khalil *et al.* [19]	Fake accounts: 13,000 Real accounts: 5,386
Twitter	Ersahin *et al.* [8]	Fake accounts: 501 Real accounts: 499
	Cresci *et al.* [18]	13,101 accounts
	Walt and Eloff [20]	223,796 accounts
	Akyon and Kalfaoglu [24]	Fake accounts: 700 Real accounts: 700
	Bharti and Pandey [33]	Real accounts: 1,103
	Narayan [34]	Fake accounts: 1,056 Real accounts: 1,176
Instagram	Meshram *et al.* [14]	Fake accounts: 3,231 Real accounts: 6,868
	Purba *et al.* [15]	Fake accounts: 32,869 Real accounts: 32,460
	Sheikhi [1]	Fake accounts: 3,132 Real accounts: 6,868
	Durga and Sudhakar [40]	Fake accounts: 201 Real accounts: 1,002

Table 3. The dataset from repositories

Social media	Reference	Dataset name	Sample size
Facebook	Elyusufi *et al.* [11]	Facebook fake profile dataset	2,816 accounts
	Reddy *et al.* [10], Shreya *et al.* [39]	Facebook profile dataset	600 accounts
Twitter	Heidari *et al.* [30]	Cresci-2017 dataset	Fake accounts: 9,262 Real accounts: 3,474
	Jabardi and Hadi [29]	The fake project dataset	11,737 accounts
	Khaled *et al.* [26]	MIB dataset	Fake accounts: 3,351 Real accounts: 1,950
	Wang *et al.* [31]	CLEF2019 dataset	7,120 accounts
	Bharti and Pandey [33]	The fake project [18]	5.870 accounts
	Chakraborty *et al.* [36]	MIB dataset	Fake accounts: 3,474 Real accounts: 3,351
	Kadam and Sharma [38]	GitHub	2,820 accounts
Instagram	Kesharwani *et al.* [32]	Fake, spammer, and genuine Instagram accounts	696 accounts
	Das *et al.* [37]	Kaggle dataset	576 accounts

Various methods are used in gathering and compiling new datasets. Some of them take advantage of third-party websites [15], web data crawlers, and social media API. After data has been gathered, commonly the fake accounts and the real accounts are separated manually. There are also other methods to simplify the data-gathering process without classifying the accounts one by one. The method used by Khalil *et al.* [19]

involved a university’s Twitter account that has a lot of followers and verifies which accounts are real or not. Meanwhile, fake accounts are obtained by buying them from a website with affordable prices.

3.2. Feature selection

According to Elyusufi *et al.* [11], the feature selection phase is a basic concept in machine learning that affects the performance of detection and classification, hence the features can provide a significant influence on the result. This phase can be done with a few techniques like using the spearman correlation test, dimensionality reduction, the markov blanket technique, and wrapper feature selection with support vector machine (SVM) [26]. Furthermore, researchers can also choose many features that could be divided into multiple classes. Furthermore, they are inserted into the model to find the best class [18]. Table 4 presents the results of the feature selection process from several studies which includes information about which features are important for training the model. The number of features varied ranging from 4 to 49 attributes. The most used feature is the features that can be obtained by the researcher without having permission from third-party software. Most of them are related to the number of followers, the number of following accounts, likes, profile pictures, status, posts, and account names.

Table 4. Feature selection for fake account classification

Reference	Features selected	Total
Gupta and Kaushal [7]	Received likes, likes, received comments, comments, tags, tag user, tags from other users, page tags, tags in comments, page tags in the comments section, tags by other users in the comments section, shared posts, wall posts, like wall posts, comments in wall posts, used applications.	17
Elyusufi *et al.* [11]	Status, followers, friends, favorites.	4
Reddy [23]	Profile ID, name, status, followers, friends, location, account creation date, shares, gender, language.	
Wang *et al.* [31]	The average of mentions, emojis, stop words, topics, links, retweets, similar posts, post length, forwarded posts, and punctuations.	10
Walt and Eloff [20]	Account age, duplicate accounts, follower and friend ratio, followers, friends, geographical location, pictures, name, profile, URL, status, groups, username.	13
Mohammad *et al.* [27]	Likes, favorites, followings, followers, location, status replies, user replies, amount of registration, hashtags, mentions, URL, profile picture, replies, shares, the status of the account that shared, status.	16
Khalil *et al.* [19]	Status, followers, followings, favorites, ratio followings, and followers, registration.	6
Jabardi and Hadi [29]	Favorites, likes, status, location, followers, account age, friend tier, reputation, friendship, registration.	10
Heidari *et al.* [30]	Followers, friends, retweets, replies, hashtags, shared URL, text, show name, user ID, neutral posts, positive posts, negative posts, number of positive posts, number of negative posts, an average of positive posts, and an average of negative posts.	16
Ersahin *et al.* [8]	Description, protected or not, followers, friends, status, favorites, public lists, verified, profile picture, contributors, affiliated profiles, affiliated profile picture, translator, hashtags, mentions, URL.	16
Cresci *et al.* [18]	Profile feature (features consist of information in the follower’s profile of the target account), timeline feature (information of tweets in the follower’s timeline of the target account), relationships feature (features from accounts that has a connectiona with the target account’s followers).	49
Sheikhi [1]	Profile picture, followed accounts, whether the follower count is greater, and the number of posts.	4
Purba *et al.* [15]	Posts, following, followers, biography, link, length of description, the presence of a description, the presence of pictures, likes, comments, location, hashtags, keywords, followers, post similarities, posts per hour.	17
Meshram *et al.* [14]	Post count, followers, followings, profile picture, private or public account, biography, username length, numbers in the username.	8
Akyon and Kalfaoglu [24]	Media number total, followers, following, numbers of integer in name, private or public account.	5
Bharti and Pandey [33]	Number of followers, friends, tweets per day, status count, mentions, and hashtags per tweet, added into a user’s favorite list, has over 50 tweets, URL, followers to the following ratio, replies.	11
Chakraborty *et al.* [36]	Number of friends, followers, status, favorites, listed count, language count, geo-enabled.	7
Durga and Sudhakar [40]	Numbers of followers, followings, media, biography count, profile picture, private account, username digit count, username length, biography emoji count.	9
Pashwan and Ravi [35]	Numbers of followers, friends, favorites, tweets, tweet frequency, location, verified account.	7
Shreya *et al.* [39]	User age, gender, account age, link in the description, status, friends count, location, location IP, status.	9

3.3. Machine learning model

Machine learning is used to perform the detection process of fake accounts on social media. The majority of research studies used more than one algorithm to find the best model. Combining 2 algorithms has been possible to increase the accuracy like a study conducted by Mohammad *et al.* [27]. They combined a convolutional neural network (CNN) and an artificial neural networks (ANN) model. Another was by Khaled *et al.* [26] which integrated an SVM with an ANN model. Table 5 lists the algorithms used in several kinds of research to create a fake account classification model according to the target social media. From Table 5, we can conclude that 38 algorithms can be used for the fake account classification task. 2 of them are a combination of 2 classification methods. According to the result, the most used method to detect fake accounts on Facebook is random forest. On the other hand, the SVM method is used commonly for Twitter. Instagram has several common approaches such as ANN, naïve bayes, random forest, and SVM.

Table 5. Classification algorithm used based on the social media platform

Social media	Algorithm	Reference	Total
Facebook	AdaBoost	[22], [25]	2
	Bagging	[25]	1
	XGBoost	[25]	1
	GradientBoost	[25]	1
	Random forest	[22], [23], [25], [39]	4
	Logistic regression	[25]	1
	Linear support vector classification (SVC)	[25]	1
	ExtraTree	[25]	1
	SVM	[7], [21], [39]	3
	Complement naïve bayes (CNB)	[21]	1
	K-nearest neighbor (KNN)	[7]	1
	ANN	[10], [39]	2
	Naïve bayes	[7], [11]	2
	Decision tree	[7], [11]	2
	Decision rule based	[7]	1
	C4.5	[22]	1
Twitter	KNN	[18], [19], [31], [38]	4
	Local outlier factor (LOF)	[31]	1
	IForest	[31]	1
	One-class SVM (OCSVM)	[31]	1
	Histogram-based outlier score (HBOS)	[31]	1
	Feature bagging	[31]	1
	Principal component analysis (PCA)	[31]	1
	Minimum covariance determinant (MCD)	[31]	1
	Variational auto encoder (VAE)	[31]	1
	Random forest	[18], [20], [28], [30], [34], [36]	6
	AdaBoost	[18], [20], [28], [36]	4
	XGBoost	[28], [36]	2
	SVM	[18]–[21], [26], [29], [30], [35], [38]	9
	CNB	[21]	1
	CNN	[27]	1
	ANN	[26], [27], [30], [38]	4
	CNN-ANN	[27]	1
	Simple logistic regression	[18], [19], [28]–[30], [33]	6
	SVM-ANN	[26]	1
	Naïve bayes	[8], [29], [33], [34], [38]	5
	Decorate	[18]	1
	Decision tree	[18], [33], [34], [36]	4
	Bayesian network	[18]	1
	Ontology	[29]	1
	Logistic with particle swarm optimization (PSO)	[33]	1
	C4.5	[38]	1
Instagram	ANN	[14], [24], [32]	3
	CNB	[21]	1
	Decision tree	[1], [15], [40]	3
	Hoeffding tree	[1]	1
	Logistic regression	[15], [24], [37], [40]	4
	Multilayer perceptron (MLP)	[1]	1
	MLP	[15]	1
	Naïve bayes	[1], [15], [24], [37]	4
	Random forest	[1], [14], [15], [37]	4
	Radial basis function (RBF)	[1]	1
	SVC	[14]	1
	SVM	[1], [21], [24], [37]	4
	KNN	[37], [40]	2

3.4. Model performance and evaluation

After a classification model was trained, an evaluation process is performed to know the performance of the model. A confusion matrix is a common evaluation method. According to Reddy [23], the confusion matrix is a technique to summarize the performance of a classification model. Using this technique, researchers can assess how well a model performs with its characteristics. The values that are generally used in this evaluation are the recall values (the ratio between true positive and the whole positive), precision values (the ratio between true positive and the results of positive detection), F1 values (the accuracy values obtained from the recall and precision values), and accuracy (how can an algorithm predict correctly). The algorithms with the highest performance from each research are presented in Table 6.

Table 6. Highest-performing models for fake account detection

Social media	References	Algorithm	Performance (%)
Facebook	Singh and Banerjee [25]	AdaBoost	Precision: 99 Recall: 99 F1 value: 99
	Raturi [21]	SVM	Accuracy: 97
	Gupta and Kaushal [7]	Not mentioned	Accuracy: 79
	Elyusufi *et al.* [11]	Decision tree (J48)	Accuracy: 99.28
	Reddy [23]	Random forest	Precision: 91 Recall: 90 F1 value: 90
	Chen and Wu [22]	Random forest	Accuracy: 98.60
	Reddy *et al.* [10]	ANN	Accuracy: 98.33
	Shreya *et al.* [39]	ANN	Accuracy: 96.73
Twitter	Wang *et al.* [31]	KNN	Precision: 93.79 Recall: 98.79
	Walt and Eloff [20]	Random forest	Accuracy: 87.11 F1 value: 49.75
	Raturi [21]	SVM	Accuracy: 99
	Mohammad *et al.* [27]	CNN-ANN	Accuracy: 99.43 Precision: 99 Recall: 99 F1 value: 99
	Khalil *et al.* [19]	K-NN	Accuracy: 98.74
	Khaled *et al.* [26]	SVM-ANN	Accuracy: 98
	Jabardi and Hadi [29]	Ontology	Accuracy: 97.2 Precision: 98.6 Recall: 97.5 F1 value: 98.0
	Heidari *et al.* [30]	Random forest	Accuracy: 89.1 F1 Value: 90.1
	Erşahin *et al.* [8]	Naïve bayes	Accuracy: 90.9
	Cresci *et al.* [18]	Random forest	Accuracy: 97.5 Precision: 98.2 Recall: 97.5 F1 value: 97.9
	Bharti and Pandey [33]	Logistic with PSO	Accuracy: 96.2 F1 value: 89
	Kadam and Sharma [38]	ANN	Accuracy: 97.4
	Chakraborty *et al.* [36]	XGBoost	Accuracy: 99.6
	Narayan [34]	Decision tree	Accuracy: 93
	Pakaya *et al.* [28]	XGBoost	Accuracy: 95.55
	Pashwan and Ravi [35]	SVM	Accuracy: 97.33
Instagram	Sheikhi [1]	Bagging decision tree	Accuracy: 98.45 Recall: 99.2 F-score: 98.9
	Raturi [21]	SVM dan CNB	Accuracy: 95
	Purba *et al.* [15]	Decision tree 2-Class	Accuracy: 79.66 Precision: 80 Recall: 79.7 F1 value: 79.6
	Meshram *et al.* [14]	Random forest	Accuracy: 96.94 Precision: 99 Recall: 98 F1 value: 98
	Kesharwani *et al.* [32]	ANN	Accuracy: 93.63
	Akyon and Kalfaoglu [24]	SVM	Precision: 91 Recall: 98 F1 value: 86
	Das *et al.* [37]	Random forest	Accuracy: 98 Recall: 97 F1 value: 98
	Durga and Sudhakar [40]	Decision tree	Precision: 96 Recall: 96 F1 value: 96

The result shows that the dominant algorithms that have the highest performance are random forest (6 articles) and SVM (3 articles). Generally, fake account detection models can be classified with an accuracy of over 90%. The highest accuracy is obtained by the decision tree classifier (99.28%) for Facebook [11], XGBoost (99.6%) for Twitter [36], and bagging decision tree (98.45%) for Instagram [1]. From this result, we can conclude that the decision tree is the best model to detect fake accounts compared to the other algorithms. Furthermore, the integration between two models can also produce a high-performance model. This approach can open opportunities to combine a variety of algorithms to increase the quality, speed, and efficiency of classification models.

4. CONCLUSION

As mentioned in the previous part, numerous researchers are aiming to create a fake account detection model using a machine learning model with a variety of algorithms. The majority of features used to train these models are gathered from data that is available from a user's profile and online activities. Using this information, researchers can identify some fake accounts that are used to commit cybercrime automatically. This literature review has found 39 algorithms that can be used as a classification model for fake account detection problems. The most used methods are random forest and SVM. From all algorithms, the combination of the decision tree and the CNN-ANN model can provide the highest performance for all three social media platforms (i.e., Facebook, Instagram, and Twitter) with an accuracy exceeding 98%.

REFERENCES

[1] S. Sheikhi, "An Efficient Method for Detection of Fake Accounts on the Instagram Platform," *Revue d'Intelligence Artificielle*, vol. 34, no. 4, pp. 429–436, Sep. 2020, doi: 10.18280/ria.340407.

[2] M. Zirui and G. Bin, "A Privacy-Preserved and User Self-Governance Blockchain-Based Framework to Combat COVID-19 Depression in Social Media," *IEEE Access*, vol. 11, pp. 35255–35280, 2023, doi: 10.1109/access.2023.3264598.

[3] D. Sharma and E. R. S. Madan, "ANN based Fake User Profile Detection," *International Journal for Research in Engineering and Emerging Trends (IJREET)*, vol. 6, no. 2, pp. 460–465, 2022.

[4] K. Krombholz, D. Merkl, and E. Weippl, "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, pp. 175–212, Dec. 2012, doi: 10.1007/s12927-012-0008-z.

[5] A. Anand, S. Dutta, and P. Mukherjee, "Influencer Marketing with Fake Followers," *IIM Bangalore Research Paper*, pp. 1–22, 2022, doi: 10.2139/ssrn.3306088.

[6] R. Nennuri, M. G. Yadav, B. Shara, G. A. Kumar, and M. Shivani, "Fake Account Detection using Machine Learning and Data Science," *Annals of the Romanian Society for Cell Biology*, pp. 6857–6865, vol. 25, no. 6, Jun. 2021.

[7] A. Gupta and R. Kaushal, "Towards detecting fake user accounts in facebook," in *2017 ISEA Asia Security and Privacy (ISEASP)*, IEEE, pp. 1-6, Jan. 2017, doi: 10.1109/iseasp.2017.7976996.

[8] B. Ersahin, O. Aktas, D. Kilinc, and C. Akyol, "Twitter fake account detection," in *2017 International Conference on Computer Science and Engineering (UBMK)*, IEEE, pp. 388-392, Oct. 2017, doi: 10.1109/ubmk.2017.8093420.

[9] S. Joshi, H. G. Nagariya, N. Dhanotiya, and S. Jain, "Identifying Fake Profile in Online Social Network," *ACI'21: Workshop on Advances in Computational Intelligence at ISIC 2021*, pp. 1–10, Feb. 2021.

[10] D. B. S. Reddy, D. R. A. Nayana, G. Sahaja, and G. S. Priya, "Artificial Intelligence Tool for Fake Account Detection from Online Social Networks," *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 14, no. 1, pp. 243-254, 2023.

[11] Y. Elyusufi, Z. Elyusufi, and M. A. Kbir, "Social networks fake profiles detection based on account setting and activity," in *Proceedings of the 4th International Conference on Smart City Applications*, ACM, pp. 1–5, Oct. 2019, doi: 10.1145/3368756.3369015.

[12] A. Alharbi, H. Dong, X. Yi, Z. Tari, and I. Khalil, "Social Media Identity Deception Detection," *ACM Computing Surveys*, vol. 54, no. 3, pp. 1–35, May 2021, doi: 10.1145/3446372.

[13] K. Nakagawa, N. T. Yang, M. Wilson, and P. Yellowlees, "Twitter Usage Among Physicians From 2016 to 2020: Algorithm Development and Longitudinal Analysis Study," *Journal of Medical Internet Research*, vol. 24, no. 9, pp. 1-10, Mar. 2022, doi: 10.2196/37752.

[14] E. P. Meshram, R. Bhambulkar, P. Pokale, K. Kharbikar, and A. Awachat, "Automatic Detection of Fake Profile Using Machine Learning on Instagram," *International Journal of Scientific Research in Science and Technology*, pp. 117–127, May 2021, doi: 10.32628/ijsrst218330.

[15] K. R. Purba, D. Asirvatham, and R. K. Murugesan, "Classification of instagram fake users using supervised machine learning algorithms," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 3, pp. 2763-2772, Jun. 2020, doi: 10.11591/ijece.v10i3.pp2763-2772.

[16] N. A. Zuhroh and N. A. Rakhmawati, "Clickbait detection: A literature review of the methods used," *Register: Jurnal Ilmiah Teknologi Sistem Informasi*, vol. 6, no. 1, pp. 1-10, Oct. 2019, doi: 10.26594/register.v6i1.1561.

[17] B. Kitchenham and S. Charters, "Guidelines for performing Systematic Literature Reviews in Software Engineering," *IEEE Access*, vol. 2, pp. 1–57, Jan. 2007.

[18] S. Cresci, R. D. Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi, "Fame for sale: Efficient detection of fake Twitter followers," *Decision Support Systems*, vol. 80, pp. 56–71, Dec. 2015, doi: 10.1016/j.dss.2015.09.003.

[19] A. Khalil, H. Hajjdiab, and N. Al-Qirim, "Detecting Fake Followers in Twitter: A Machine Learning Approach," *International Journal of Machine Learning and Computing*, vol. 7, no. 6, pp. 198–202, Dec. 2017, doi: 10.18178/ijmlc.2017.7.6.646.

[20] E. V. D. Walt and J. Eloff, "Using Machine Learning to Detect Fake Identities: Bots vs Humans," *IEEE Access*, vol. 6, pp. 6540–6549, 2018, doi: 10.1109/access.2018.2796018.

[21] R. Raturi, "Machine Learning Implementation for Identifying Fake Accounts in Social Network," *International Journal of Pure and Applied Mathematics*, vol. 118, no. 20, pp. 4785–4797, Aug. 2018.

[22] Y. -C. Chen and S. F. Wu, "FakeBuster: A Robust Fake Account Detection by Activity Analysis," in *2018 9th International Symposium on Parallel Architectures, Algorithms and Programming (PAAP)*, IEEE, pp. 108-110, Dec. 2018, doi: 10.1109/paap.2018.00026.

[23] S. D. P. Reddy, “Fake Profile Identification using Machine Learning,” *International Research Journal of Engineering and Technology (IRJET)*, vol. 6, no. 12, pp. 1145–1150, 2019.

[24] F. C. Akyon and M. E. Kalfaoglu, “Instagram Fake and Automated Account Detection,” in *2019 Innovations in Intelligent Systems and Applications Conference (ASYU)*, IEEE, pp. 1-7, Oct. 2019, doi: 10.1109/asyu48272.2019.8946437.

[25] Y. Singh and S. Banerjee, “Fake (Sybil) Account Detection Using Machine Learning,” *SSRN Electronic Journal*, pp. 1-7, 2019, doi: 10.2139/ssrn.3462933.

[26] S. Khaled, N. El-Tazi, and H. M. O. Mokhtar, “Detecting Fake Accounts on Social Media,” in *2018 IEEE International Conference on Big Data (Big Data)*, IEEE, pp. 3672-3681, Dec. 2018, doi: 10.1109/bigdata.2018.8621913.

[27] S. Mohammad, M. U. S. Khan, M. Ali, L. Liu, M. Shardlow, and R. Nawaz, “Bot detection using a single post on social media,” in *2019 Third World Conference on Smart Trends in Systems Security and Sustainablity (WorldS4)*, IEEE, pp. 215-220, Jul. 2019, doi: 10.1109/worlds4.2019.8903989.

[28] F. N. Pakaya, M. O. Ibrohim, and I. Budi, “Malicious Account Detection on Twitter Based on Tweet Account Features using Machine Learning,” in *2019 Fourth International Conference on Informatics and Computing (ICIC)*, IEEE, pp. 1-5, Oct. 2019, doi: 10.1109/icic47613.2019.8985840.

[29] M. Jabardi and A. S. Hadi, “Twitter Fake Account Detection and Classification using Ontological Engineering and Semantic Web Rule Language,” *Karbala International Journal of Modern Science*, vol. 6, no. 4, pp. 403-413, Dec. 2020, doi: 10.33640/2405-609x.2285.

[30] M. Heidari, J. H. J. Jones, and O. Uzuner, “An Empirical Study of Machine learning Algorithms for Social Media Bot Detection,” in *2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*, IEEE, pp. 1-5, Apr. 2021, doi: 10.1109/iemtronics52119.2021.9422605.

[31] X. Wang, Q. Zheng, K. Zheng, Y. Sui, S. Cao, and Y. Shi, “Detecting Social Media Bots with Variational AutoEncoder and k-Nearest Neighbor,” *Applied Sciences*, vol. 11, no. 12, pp. 1-15, Jun. 2021, doi: 10.3390/app11125482.

[32] M. Kesharwani, S. Kumari, and V. Niranjan, “Detecting Fake Social Media Account Using Deep Neural Networking,” *International Research Journal of Engineering and Technology (IRJET)*, vol. 8, no. 7, pp. 1191–1197, Jul. 2021.

[33] K. K. Bharti and S. Pandey, “Fake account detection in twitter using logistic regression with particle swarm optimization,” *Soft Computing*, vol. 25, no. 16, pp. 11333–11345, Jun. 2021, doi: 10.1007/s00500-021-05930-y.

[34] N. Narayan, “Twitter Bot Detection using Machine Learning Algorithms,” in *2021 Fourth International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, IEEE, pp. 1-4, Sep. 2021, doi: 10.1109/icecct52121.2021.9616841.

[35] V. A. Pashwan and D. S. Ravi, “Fake Profile Detection on Twitter using SVM Classifier,” *Indian Journal of Signal Processing (IJSP)*, vol. 1, no. 1, pp. 16–20, Feb. 2021.

[36] P. Chakraborty, M. M. Shazan, M. Nahid, M. K. Ahmed, and P. C. Talukder, “Fake Profile Detection Using Machine Learning Techniques,” *Journal of Computer and Communications*, vol. 10, no. 10, pp. 74–87, 2022, doi: 10.4236/jcc.2022.1010006.

[37] S. Das, S. Saha, S. Vijayalakshmi, and J. Jaiswal, “An Effecient Approach to Detect Fraud Instagram Accounts Using Supervised ML Algorithms,” in *2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, IEEE, pp. 760-764, Dec. 2022, doi: 10.1109/icac3n56670.2022.10074364.

[38] N. Kadam and S. K. Sharma, “Social Media Fake Profile Detection Using Data Mining Technique,” *Journal of Advances in Information Technology*, vol. 13, no. 5, pp. 518-523, 2022, doi: 10.12720/jait.13.5.518-523.

[39] K. Shreya, A. Kothapelly, D. V, and H. Shanmugasundaram, “Identification of Fake accounts in social media using machine learning,” in *2022 Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT)*, IEEE, pp. 1-4, Dec. 2022, doi: 10.1109/icerect56837.2022.10060194.

[40] P. Durga and D. T. Sudhakar, “The use of supervised machine learning classifiers for the detection of fake instagram accounts,” *Journal of Pharmaceutical Negative Results*, pp. 267–279, Jan. 2023, doi: 10.47750/pnr.2023.14.03.36.

BIOGRAPHIES OF AUTHORS

Nalia Graciella Kerrysa is a bachelor's degree student in the Data Science Technology Study Program, Faculty of Advanced Technology and Multidiscipline, Universitas Airlangga, Indonesia. She is one of the active students with a passion for research. She can be contacted at email: naliagk.hulu@gmail.com.

Ika Qutsiati Utami is a doctoral degree student at the Graduate Institute of Network Learning Technology, National Central University, ROC, Taiwan. She also received an M.S. degree from the Graduate Institute of Network Learning Technology, National Central University, Taiwan, in 2018. She is currently working as a lecturer at the Faculty of Advanced Technology and Multidiscipline, Universitas Airlangga, Indonesia. Her research interests are in the areas of human-computer interaction, computer science in education, and AI/ML for future education. She can be contacted at email: ika.q.utami@ftmm.unair.ac.id.