e-ISSN: 2686-2573
DOI: 10.21512/emacsjournal.v4i1.8076

A Systematic Literature Review: Instagram Fake Account Detection Based on Machine Learning

Juandreas Ezarfelix[1], Nathanael Jeffrey[2], Novita Sari[3*]

[1,2,3] Computer Science Department, School of Computer Science,
Bina Nusantara University,
Jakarta, Indonesia 11480
juandreas.ezarfelix@binus.ac.id; nathanael.jeffrey001@binus.ac.id; novita.sari005@binus.ac.id;

*Correspondence: novita.sari005@binus.ac.id

Abstract *— The popularity of social media continues to grow, and its dominance of the entire world has become one of the aspects of modern life that cannot be ignored. The rapid growth of social media has resulted in the emergence of ecosystem problems. Hate speech, fraud, fake news, and a slew of other issues are becoming un-stoppable. With over 1.7 billion fake accounts on social media, the losses have al-ready been significant, and removing these accounts will take a long time. Due to the growing number of Instagram users, the need for identifying fake accounts on social media, specifically in Instagram, is increasing. Because this process takes a long time if done manually by humans, we can now use machine learning to identify fake accounts thanks to the rapid development of machine learning. We can detect fake accounts on Instagram using machine learning by implementing the combination of image detection and natural language processing.*

Keywords: *Fake Accounts; Social Media; Instagram; Machine Learning; Natural Language Processing; Image Detection.*

I. INTRODUCTION

Social media has become ingrained in our daily lives. Along with the rapid advancement of technology, social media is rapidly expanding and becoming accessible to anyone, anywhere [1][2]. We can interact with people from all over the world through social media without having to meet in person. Furthermore, there are numerous advantages to using social media, such as viewing news, searching for information, promoting a business, and making new friends. But, in addition to all of these advantages, social media also has a slew of draw-backs that are difficult to ignore given the scope of the medium.

Instagram is currently one of the most popular social media platforms[3]. We can share photos, videos, and our daily lives with others via Instagram. However, as previously stated, there are a slew of negative aspects to Instagram, including fraud, hate speech, and other potentially harmful activities. This is usually done by fake accounts, whether they are bot accounts or people who make me fake to conceal their identity[4].

The growing number of fake accounts that have a negative impact on social media, including Instagram, must be taken into account and investigated[4]. However, deactivating these accounts one by one would be impossible due to the large number of them. Machine learning can already be used and imple-mented in many applications and software that are useful for assisting humans in tasks or roles that machines

should be able to do due to technological advancements[5]. We also want to use or implement this technology on Insta-gram to detect fake accounts so that they can be dealt with more easily.

Image Recognition and Natural Language Processing are two areas of machine learning that can be used here[5][6]. Where both have distinct roles that can be used to identify fake Instagram accounts. We can see if the image posted by the account contains hate speech or fake news using Image Recognition. Meanwhile, we can use Natural Language Processing to identify the results of existing accounts' comments and captions to see if they contain hate speech[5][6].

We can save time and resources by using machine learning to detect fake accounts, while still getting excellent results. This detection feature can then be integrated into other social media software, such as Facebook, to help reduce the number of fake accounts on the platform.

The limitations will remain the same, namely that fake accounts will not be deleted because they were created by others or bots, so it is unlikely that we will be able to eliminate all fake accounts. We hope that by conducting this literature review, we will be able to take advantage of certain high-accuracy machine learning algorithms to reduce the number of fake accounts on Instagram, with the hope that it will later be developed for other platforms.

1.1 Literature Review

This paper was put together by combining data from previous studies.

Table I. State of The Art of Fake Account In Social Media.

No.	Title	Year	Method	Sample	Evaluation
1	Prediction of Fake Instagram Profiles Using Machine Learning	2021	Using the combination of image detection and Natural Language Processing (NLP) to detect fake accounts on Instagram	Web scrapping datasets from instagram that's being labelled for the training	The results shows that using image detection such as CNN and combine with NLP results in a 91.5% accuracy of fake account detection
2	Instagram Fake and Automated Account Detection	2019	Detection of accounts using several Machine Learning algorithms such as Naïve Bayes, Logistic Regression, SVM, and Neural Networks	2 Datasets consisting of 700 real account and 700 automated account gathered from different countries and fields	The result shows Neural Network works the best for identifying fake account while SVM works the best for identifying automated account
3	Automatic Detection of Fake Profile Using Machine Learning on Instagram	2021	Using several Machine Learning algorithm to detect fake accounts such as ANN, Random Forest, and SVC	1002 actual instagram accounts and 201 fake accounts has been collected for labeling	Artificial Neural Netowork shows the best result in detecting fake accounts because of the nature of the datasets which are images
4	Classification of instagram fake users using supervised machine learning algorithms	2020	Using several Machine Learning algorithm to detect instagram fake accounts such as Random Forest, Multilayer Perceptron, Logistic Regression, Naives Bayes, and J48 Decision Tree	Data from web scrapping from instagram websites with the total of 32.460 users to be used as a dataset	The result from the research shows that Random Forest and J48 Decision Tree give the best accuracy in terms of detecting fake accounts from instagram
5	An efficient method for detection of fake accounts on the instagram platform	2020	Image classification using Machine Learning to identify fake accoutns	10.000 Instagram accounts as experiments	The result is the model got 90%+ accuracy on detecting fake accounts from 10.000 instagram accounts
6	Using Machine Learning to Detect Fake Identities: Bots vs Humans	2018	Using supervised machine learning models to train the datasets (SVM, rf, and Adaboost)	Datasets from the previouse researchers datasets like paedophiles and extremism groups	The current machine learning models and datasets are not suited to detect bot accounts in social media
7	Detection of Fake Accounts in Instagram Using Machine Learning	2019	Using Machine Learning techniques that are Logistic Regression and Random Forest Algorithm	Kaggle datasets that consist of fake and legitimate datasets	The accuracy of using Logistic Regression is 90.8% while Random forest algorithm got 92.5%
8	Fake account detection using machine learning and data science	2019	Using Decision tree and Gradient Boosting as the new more effective machine learning algorithm for classification problem	Data from web scraper to extract necessary information from social media such as login activity, likes, comments, number of posts, followings, and followers	The results from the new algorithm resulting in a higher accurate results in detecting fake accounts
9	Detecting Fake Social Media Account Using Deep Neural Networking	2021	Six Layers of Artificial Neural Network is used to train the datasets in order to detect fake instagram accounts	696 instagram accounts where there are half real and half fake accounts	The model gives a 93.63% accuracy in detecting fake accounts with a 0.18% loss
10	Machine Learning Implementation for Identifying Fake Accounts in Social Network	2018	The combination of NLP and network identification to identify the account details based on the networks and SVM BOW concept for identification of the number of words are harmful	15.000 of each the following social media: Facebook, Instagram, Twitter, Youtube, Whatsapp	Using the two methods metioned, around 20% accounts for each social media detected as invalid acconts of which some get deleted

1.2 Social Media

Social media is a computer-based technology that allows people to share their ideas, opinions, and information through virtual networks and communities. Social media is an internet-based platform that allows people to share content such as personal information, documents, films, and images quickly and electronically. Users interact with social media using web-based software or applications on a computer, tablet, or smartphone. While social media is widely used in the United States and Europe, Asian countries such as Indonesia are at the top of the list. As of October 2021, around 4.5 billion people utilize social media.[8]

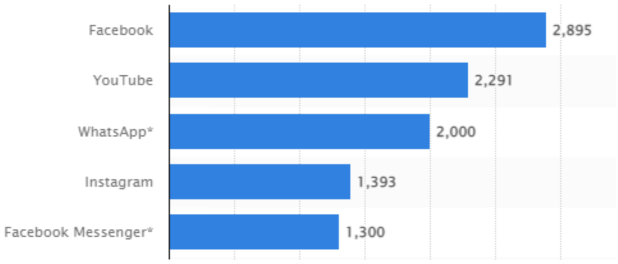

Fig. 1. Most popular Social Networks Worldwide Oct 2021

"Most popular social networks worldwide as of October 2021, ranked by number of active users," according to this data. Facebook, the most popular social media platform, was the first to cross one billion registered accounts, and it now has more than 2.89 billion monthly active users. Facebook (main platform), WhatsApp, Facebook Messenger, and Instagram are the company's four largest social media platforms, each with over one billion monthly active users. Facebook claimed over 3.58 billion monthly core Family product users in the third quarter of 2021. [8]

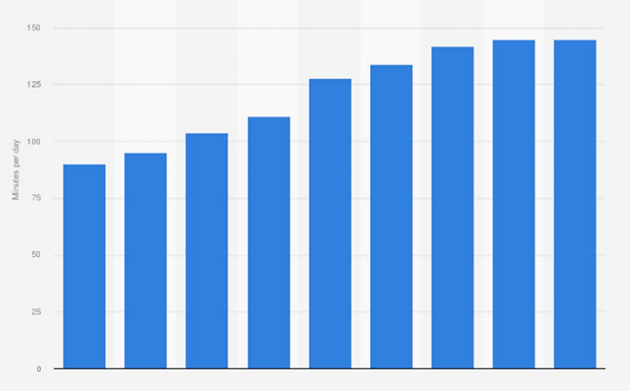

Fig. 2. Daily time spent on social networking by internet users worldwide from 2012 to 2020

In 2019 and 2020, internet users worldwide spent an average of 145 minutes per day on social media, up from 142 minutes the previous year. The Philippines currently has the highest time spent on social media per day, with online users spending an average of three hours and 53 minutes each day on the platform. In comparison, Americans spend only two hours and three minutes every day on social media. [8] They used social media to stay in touch with friends and family, filling spare time with reading news in social media, and to follow celebrities or influencers.[9]

Social media has a huge impact on not just online activity, but also offline behavior and everyday life. In a global online user study conducted in February 2019, a large majority of respondents said that social media had improved their access to information, communication, and freedom of speech. On the other hand, respondents said that social media had harmed their personal privacy, increased political polarization, and increased daily diversions. [2]

1.3 Fake Account

Fake accounts are those that are created in order to boost the popularity of other users. As a result, they tend to have a large following and a small number of followers. Their preferences may appear to be random. Fake accounts are characterized by the absence of a profile picture and an unusual username. [11]

Fake accounts are risky for social media platforms because they can change notions like popularity and influence on Instagram, as well as have an impact on the economics, politics, and society. For the Instagram platform, this research proposed a machine learning-based fake account detection method.

Many individuals create fake accounts for a variety of reasons, including hate speech, fraud, fake news, and a variety of other difficulties. It may have an impact on an account's social media reputation and Instagram insight. These methods include the use of bots, the purchase of social metrics such as likes, comments, and followers, and the use of platforms or networks that allow users to exchange metrics.

Fake accounts proliferating on social media have resulted in a flood of uncontrollable fake news, which is extremely harmful for celebrities. Some artists are in trouble as a result of fake news spreading on Instagram. They must clarify by reporting fake news that defames their name to the authorities. This fake account can be a way for the cyberbullying; real users also have different anxieties about their privacy in the online environment with these fake accounts

Therefore, over the past years, many researchers have investigated the problem of detecting malicious activities and spammers in social media using machine learning techniques. However, there are a limited number of research articles relating to detecting fake accounts. [17]

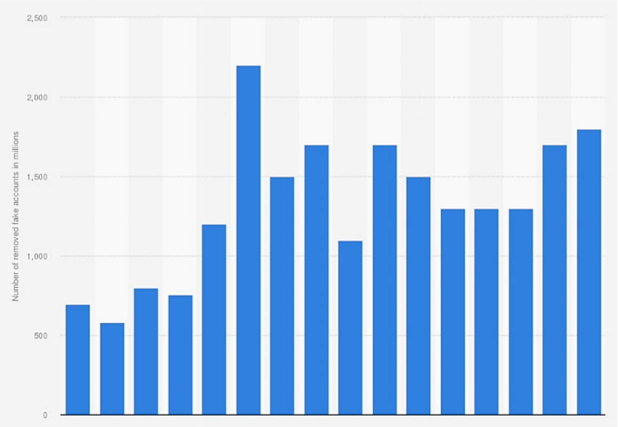

Fig. 3. Global number of fake accounts taken action on by Facebook from 4th quarter 2017 to 3rd quarter 2021

From this statistic we can see the number of fake accounts worldwide that Facebook delete from 4th quarter 2017 to 3rd 2021. Recently facebook, delete approximately 1.8 billion fake accounts, up from 1.3 billion fake accounts in same quarter in 2020. [10]

II. METHOD

The method used for this paper is a systematic literature review. First, we make the research questions based on our topics. Next, we collect some research papers and design the state of the art. In order to evaluate our paper, we used PRISMA checklist to help us review our paper in this study. The workflow of our study can be seen in the diagram below.

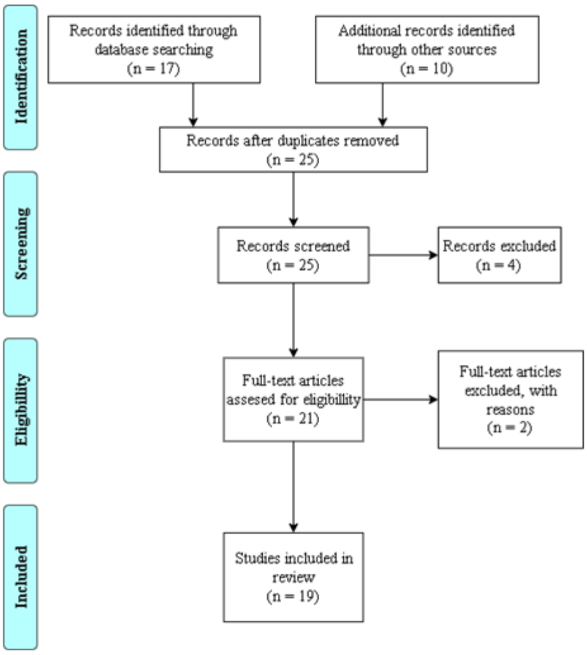
Identification
Screening
Eligibillity
Included
Records identified through database searching (n = 17)
Additional records identified through other sources (n = 10)
Records after duplicates removed (n = 25)
Records screened (n = 25)
Records excluded (n = 4)
Full-text articles assesed for eligibillity (n = 21)
Full-text articles excluded, with reasons (n = 2)
Studies included in review (n = 19)

Fig. 4. Diagram design the state of the art

The systematic review are being used to study more about the used of machine learning in detecting fake accounts in Intagram. With that in mind we have made these research question to further support this paper :

1. RQ1: What is the impact of fake accounts circulating on social media?
2. RQ2: How can machine learning help identify fake accounts on social media?
3. RQ3: What is the most efficient machine learning method to use to identify fake accounts on social media?

The main goal of this research is to look over and assess the findings of previous studies to see if machine learning can help identify fake accounts on social media platforms like Instagram.

Data is available on social media sites like Instagram that can be accessed by the general public. This data is known as metadata, and it is through metadata that we can obtain information that has been widely used in previous research. The following are some examples:

- The number of followers on an account
- The number of people who follow a particular account
- An account's total number of posts
- An account's use of a large number of hashtags
- The account's comments and likes
- An account's profile picture

Previous research has used this available metadata to analyze whether an Instagram account is fake or not using machine learning.

2.1 Datasets

Several previous studies used datasets that are made up of publicly available metadata. Where the dataset is derived from the results of personal data scraping [11] [14], as well as datasets created by others. [5] [12][13].

The datasets collected usually fall into two categories: real account datasets and fake account datasets, both of which contain the same information but with different values.

The results of grouping datasets can also be used to create a feature that shows the difference between real and fake accounts.

For example, here's a comparison of real accounts and fake accounts based on the number of followers and the number of followers [11]

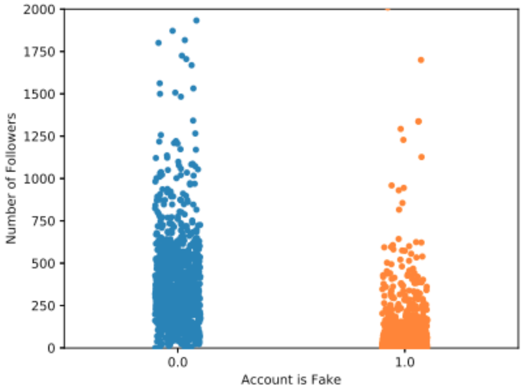

Fig. 5. Data distributions of follower counts between real account (0.0) and fake account (1.0)

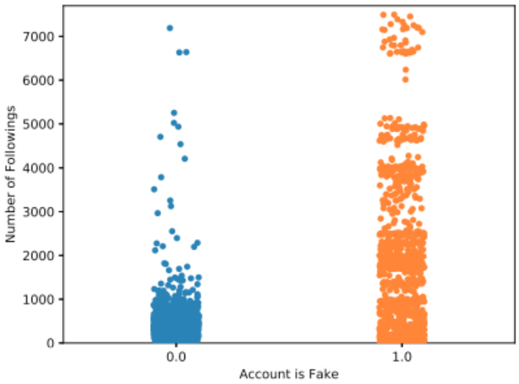

Fig. 6. Data distributions of following counts between real account (0.0) and fake account (1.0)

2.2 Proposed Method

The method that is frequently used, based on the results of several previous studies, is to use machine learning algorithm methods such as:

- Logistics Regression
- Naive Bayes
- Random Forest
- Support Vector Machine

The available metadata can be processed using the algorithms above, and the results can be turned into a reference that can be used to predict fake accounts on social media.

There are also those who use computer vision algorithms, such as neural networks, to identify social media accounts based on images, in addition to machine learning algorithm methods. Whether it's a profile photo, images posted by the account, images liked by the account, or images on the account tag, images are important.

2.2.1 Logistic Regression

The logistic regression method makes use of the logit function to predict results, with a larger result equal to the decision limit belonging to a fake account and a smaller result equal to the decision limit belonging to a real account.

By comparing the predicted results of correct fake accounts with the real number, then comparing the predicted results of false accounts with the real number, and finally comparing the predicted number of real accounts with the real number, the level of prediction accuracy from logistic regression can be calculated. The level of accuracy generated by logistic regression on the datasets used is the end result.[12]

2.2.2 Neural Network

Neural networks are a type of machine learning that is based on the way the human brain functions [5]. When looking at an object, neural networks are designed to follow the human brain's search for work, in which the brain sends signals from one neuron to another [16].

The Artificial Neural Network is a type of neural network that is frequently used (ANN). Starting with the Input Layer, the ANN is made up of many layers.

The Input Layer is the first neuron layer in the ANN, and its contents are the first data that the system receives. The ANN process will always start with the input layer.

Hidden layers are the layers in an ANN that come after the input layer and apply weights to the inputs before directing them through an activation function as the output. The mathematical functions that make up the hidden layers are used to generate a probability [17].

The Output Layer is the ANN's final layer, where the output is issued in the form of the final results of the ANN's predictions.

III. RESULT

Table I. Logistic regression algorithm average accuracy from recent research on fake account detection

Ref, Year	Accuracy
[19], 2021	90,83%,
[12], 2021	90,8%,
Average	90,815%

Notes: Different datasets and features can be used by the compared methods.

Table 2. Naïve Bayes algorithm average accuracy from recent research on fake account detection

Ref, Year	Accuracy
[17], 2020	94,58%
Average	94,58%

Notes: Different datasets and features can be used by the compared methods.

Table 3. Random Forest algorithm average accuracy from recent research on fake account detection

Ref, Year	Accuracy
[17], 2020	97,2%
[19], 2021	94,16%
[13], 2021	96,94%
[12], 2021	92,5%
Average	95,2%

Notes: Different datasets and features can be used by the compared methods.

Table 4. Support Vector Machine algorithm average accuracy from recent research on fake account detection

Ref, Year	Accuracy
[17], 2020	68,68%,
[11], 2019	86%
[13], 2021	86,63%
Average	80,43%

Notes: Different datasets and features can be used by the compared methods.

Table 5. Neural Network algorithm average accuracy from recent research on fake account detection

Ref, Year	Accuracy
[5], 2021	93,63%
[11], 2019	95%
[13], 2021	95.54%
Average	94,72%

Notes: Different datasets and features can be used by the compared methods.

Table 6. Summary of Machine Learning algorithm average accuracy from recent research on fake account detection

Method	Accuracy
Logistic Regression	90,815%
Naïve Bayes	94,58%
Random Forest	95.2%
Support Vector Machine	80,43%
Neural Network	94,72%

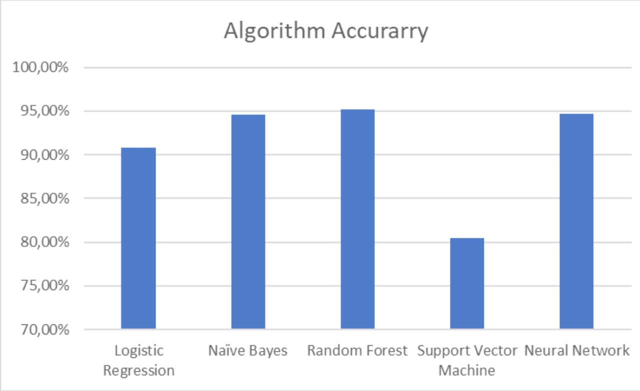

Fig. 7. Data of Machine Learning algorithm average accuracy from recent research on fake account detection

IV. CONCLUSION

Fake accounts are a concern for social media platforms because they can distort perceptions of popularity and influence on social media, as well as have economic, political, and societal consequences. For many online social media services, such as Facebook and Instagram, detecting fake accounts on social media has become a time-consuming process. This paper provides an explanation of various machine learning algorithms for detecting fake accounts on social media platforms, as well as the results of our analysis of these algorithms using data from the literature to determine which one is the most effective.As a result of a multitude of analyses, evaluations, and efforts, it has been revealed that using neural network is the most effective method to detect fake accounts.

For future research, we hope to redevelop existing machine learning algorithms by incorporating more available resources, such as object detection on neural networks, so that we can detect images on Instagram in greater detail. Along with improving accuracy, we must also improve the efficiency of existing machine learning so that it can be integrated directly into existing social media software, eliminating the need for a third-party app to detect fake accounts. With slight modifications, the machine learning method stated in this paper may also be applied for other social networking sites such as LinkedIn.

Hopefully, this paper will aid future research in reducing the number of fake accounts currently active on social media.

REFERENCES

[1]“Social media - Statistics & Facts.” *https://www.statista.com/topics/1164/social-networks/#dossierKey-figures.*

[2]“Daily time spent on social networking by internet users worldwide from 2012 to 2020”. *https://www.statista.com/statistics/433871/daily-social-media-usage-worldwide/.*

[3]“Number of social network users worldwide from 2017 to 2025”. *https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/.*

[4]“Global number of fake accounts taken action on by Facebook from 4th quarter 2017 to 2nd quarter 2021”. *https://www.statista.com/statistics/1013474/facebook-fake-account-removal-quarter/*

[5]Kesharwani, M., Kumari, S., & Niranjan, V. (2021). Detecting Fake Social Media Account Using Deep Neural Networking. July, 1191–1197.

[6]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*

[7]“Social Media Definition”. *https://www.investopedia.com/terms/s/social-media.asp*

[8]“Most popular social networks worldwide as of October 2021, ranked by number of active users”. *https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/*

[9]“Most popular reasons for internet users wordwide to use social media as of 2nd quarter 2021”. *https://www.statista.com/statistics/715449/social-media-usage-reasons-worldwide/*

[10]“Global number of fake accounts taken action on by Facebook from 4th quarter 2017 to 3rd quarter 2021” *https://www.statista.com/statistics/1013474/facebook-fake-account-removal-quarter/*

[11]Akyon, F. C., & Esat Kalfaoglu, M. (2019). Instagram Fake and Automated Account Detection. Proceedings - 2019 Innovations in Intelligent Systems and Applications Conference, ASYU 2019. *https://doi.org/10.1109/ASYU48272.2019.8946437*

[12] Dey, A., Reddy, H., Dey, M., & Sinha, N. (2019). Detection of Fake Accounts in Instagram Using Machine Learning. *International Journal of Computer Science and Information Technology*, *11*(5), 83–90. *https://doi.org/10.5121/ijcsit.2019.11507*

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

[14]Purba, K. R., Asirvatham, D., & Murugesan, R. K. (2020). Classification of instagram fake users using supervised machine learning algorithms. *International Journal of Electrical and Computer Engineering*, *10*(3), 2763–2772. *https://doi.org/10.11591/ijece.v10i3.pp2763-2772*

[15]Saranya Shree, S., Subhiksha, C., & Subhashini, R. (2021). Prediction of Fake Instagram Profiles Using Machine Learning. *SSRN Electronic Journal*. *https://doi.org/10.2139/ssrn.3802584*

[16]"Neural Network". *https://www.ibm.com/cloud/learn/neural-networks*

[17]Sheikhi, S., 2020. An Efficient Method for Detection of Fake Accounts on the Instagram Platform. Revue d'Intelligence Artificielle, 34(4), pp.429-436.

[18]"Hidden Layer Definition" *https://deepai.org/machine-learning-glossary-and-terms/hidden-layer-machine-learning*

[19]M, Mamatha, M.Srinivasa Datta, Umme Hani Ansari, Dr. Subhani Shaik. (2021). Fake Profile Identification using Machine Learning Algorithms. July, 2248-9622.