IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Real Time Detection of Fake Instagram Accounts using Selenium Automation

Pooja D Begte
BTech.(Computer Engineering) Usha Mittal Institute of Technology SNDT Women's University Mumbai

Bhavana Lilaka
BTech. (Computer Engineering) Usha Mittal Institute of Technology SNDT Women's University Mumbai

Aayushi Pandit
BTech.(Computer Engineering) Usha Mittal Institute of Technology SNDT Women's University Mumbai

Dr. Rachana Dhannawat
H.O.D. (Compueter Engineering)
Usha Mittal Institute of Technology SNDT Women's University Mumbai

***Abstract*—**Fake accounts are one of the significant problems on Instagram, which is used to spam users with unsolicited messages, advertisements, or links to malicious websites. It may be involved in various scams and fraudulent schemes, such as fake giveaways, investment scams, or online dating scams. And it may also be used to harass, intimidate, or bully individuals online. Such accounts directly affect the user's trust and enable a variety of harmful activities, including misinformation and fraud. Our research overcame all these problems and introduced a unique combination of web scraping and Selenium automation, as well as a fraud score generator. Users can input Instagram usernames, and the application gives profile data, including profile pictures, post counts, follower counts, and following counts. The algorithm initialized rate user function to assign the weights to these parameters and display the rating and comment variables. Based on the fraud score, it will categorize a given account into fake, real, and warning. This project serves as a valuable tool for users seeking to assess the credibility of Instagram accounts, providing a strong structure for authentication, analysis, and evaluation.

Index Terms—Instagram, fake account, Selenium automation, web scraping, real-time identification, fraud score generator, comment variable, parameter weightage.

I. INTRODUCTION

In today's digital age, social media platforms have emerged as essential channels for communication, networking, and self-expression. As the most popular platform, Instagram stands out as a prominent hub where millions of users share their experiences, thoughts, and creations through photos and videos. However, among the genuine users, the fake or fraudulent account challenges Instagram's integrity and user trust. Addressing the issue of fake account detection on Instagram is necessary for safeguarding the platform's credibility, encouraging a safer online environment, and preserving the trust of users. Traditional methods of manually identifying fake accounts on Instagram involve people checking each account one by one to see if it's real or fake. This process is slow and not very effective because there are so many accounts on Instagram, and it takes a lot of time to check each one. One reason why this process is hard is because fake accounts can look a lot like real ones. They might use pictures of real people, post things that seem normal, and interact with other users just like real accounts do. This makes it tricky for people to tell the difference between real and fake accounts. Another problem with manual checking is that it's a really time-consuming process. Each account has lots of things to look at, like its profile info, how many posts it has, how many followers it has, and how many people it follows.

Going through all this information for every account takes a long time, especially with so many accounts out there. Also, people can make mistakes when they're checking accounts. What one person thinks is a fake account, another person might think is real. This means there can be errors where real accounts get labeled as fake or fake accounts fall through as real ones. As Instagram grows and more people join, it becomes even harder to keep up with checking all the accounts manually. There are just too many accounts to look at, and not enough people to do the checking. In light of these concerns, there is an urgent need for effective strategies and technologies to prevent the frequency of fake accounts on Instagram. Finding out which profiles are fake is a big challenge. This study talks about using a tool named Selenium to help spot fake profiles on Instagram quickly. Selenium acts like a real person on Instagram. It can look at profiles, scroll through posts, and watch how people interact.

II. LITERATURE SURVEY

Research on detecting fake Instagram accounts has advanced over the years. Also, there is much previous research done on the detection of fake Instagram accounts. For our research, we studied many previous papers. [1] presented advanced research utilizing artificial neural networks to detect fake and automated accounts on Instagram. Using two distinct datasets, researchers achieved impressive accuracy rates, with the most successful model reaching an F1 score of 95%. This

study underscored the efficacy of deep learning techniques in discerning inauthentic activity by analyzing behavioral patterns, follower metrics, and profile information.

After that, other researchers have tried different ways to find fake accounts on Instagram. [2] explored the detection of fake accounts on Instagram using logistic regression and random forest algorithms. While both studies demonstrated the efficacy of machine learning in identifying fake accounts, they relied on pre-existing datasets rather than real-time data collection from the platform. Similarly, [3] introduced a comprehensive framework incorporating gradient-boosting algorithms. This method not only applied advanced machine learning techniques but also incorporated a range of supportive features such as web scraping, engagement rate calculations, artificial activity detection, spam comment detection, and decision trees. This paper offers a multifaceted approach to detecting fake Instagram accounts, taking into account variables like engagement rates, artificial activity, and spam comments, which are crucial indicators of authenticity on social media platforms.

Similarly, [4] presented a method for identifying and classifying fake Instagram profiles. By analyzing aspects such as account privacy, username length, and profile characteristics, the study achieved a remarkable 100% accuracy rate, showcasing the depth of analysis required for effective fake account detection. Similarly, [5] focused on account behavior, posting patterns, and follower count in detecting fake Instagram accounts, and they achieved an accuracy rate of 83.5%. In July 2022, the International Journal of Research in Engineering, IT, and Social Sciences (IJREISS) presented a study on detecting fake Twitter accounts and bots by using a support vector machine and neural network, and they achieved 91% accuracy. The method focused on recognizing patterns in account behavior and interactions, as well as profile information, which is key to identifying suspicious activity indicative of fake accounts on social media.

[6] discussed various methodologies for detecting fake accounts on social media through machine learning. The algorithms utilized in this study included support vector machines (SVM), artificial neural networks (ANN), decision trees, K-nearest neighbors (KNN), and naive bayes. With a focus on feature selection and classification, the paper delves into the application of these algorithms to effectively identify fake profiles, which is a significant contribution to ensuring the integrity and trustworthiness of social media ecosystems.The previous methods mainly used machine learning algorithms to spot fake accounts on Instagram. Also they used Dummy dataset for detection of fake accounts. But we are introducing the recent advancements in Instagram fake account detection. Instead of using a specific algorithm, we are using a new approach called selenium. Selenium is a tool for real-time data collection and analysis directly from the Instagram platform. By examining parameters such as profile picture, post count, followers versus following ratio, and profile description, this approach achieved an accuracy rate of 91% across 220 profiles. This contrasts with previous studies, which primarily focused on algorithmic approaches. It shows that directly interacting with Instagram as it happens can be a great way to find fake accounts quickly and accurately.

III. METHODOLOGY

The proposed model aims to improve the existing system in several ways to make it more user-friendly, secure, and efficient. Firstly, we plan to enhance user management features to better organize users and their permissions. This means making sure only authorized users can access certain parts of the system. Next, we want to make sure we're getting our data from Instagram in the best way possible. Right now, we're using a method called web scraping, which is like copying and pasting information from a webpage. For this purpose, we used Selenium, which gives automation to the system and retrieves the data from the instagarm. system provides a secure and automated login process for the user. The application uses authentication system for user login and logout functionalities. In the user session, we created "admin login," "logoutUser," and "home" for managing access control.

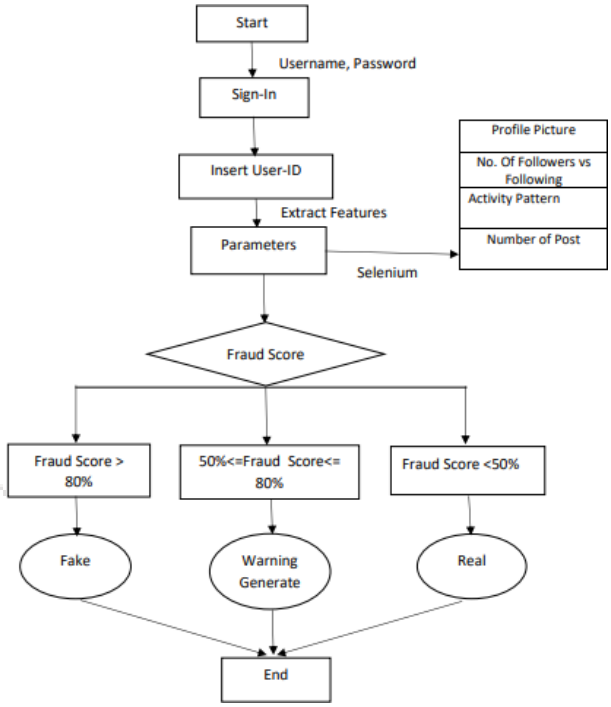

Fig. 1. Flowchart.

This flowchart shows a step-by-step plan to determine if a social media post is real or fake by giving it a lie score. First, you log in using your username and password. Then enter the social media ID of the profile you want to check. The system looks at different information about the profile, such as the picture of the profile, the number and number of people it follows, the content of the profile, the username, how it works and how much money it has. This content helps calculate the false score as a percentage

Information based on the false score is three. It will be marked in one of the following ways: If the score is above 80%, the profile is considered fake. If the score is between 50% and 80%, you will receive a warning that it may not be accurate. If the score is below 50%, the profile is considered correct. This method will help you quickly identify which social media profiles may not be legitimate.

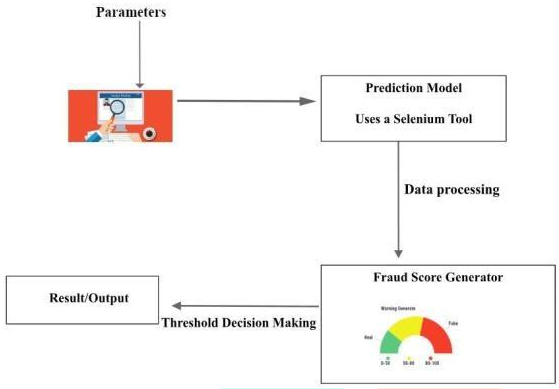

Fig. 2. Block Diagram.

This block diagram gives a detailed view of our system. It contains dataset collection, a fake account detection algorithm, fraud score calculation, a comment variable, and data storage & display. The steps are explained in detail below.

A. Dataset Collection

In our research, we gathered real-time user accounts from Instagram, with a specific focus on business accounts. This dataset contains fake, real and suspicious accounts.To compile our dataset, we manually created a collection of business accounts from Instagram, ensuring a diverse representation across different industries, sectors, and geographic regions. The dataset obtained from this process is structured, meaning that it is organized in a systematic manner, facilitating efficient analysis and interpretation. Each entry in the dataset contains essential information about a specific Instagram business account, including account information, engagement metrics, content analysis, and metadata extracted from Instagram user profiles, particularly focusing on business accounts. The account information section includes essential details such as the username, display name, profile picture, bio, and website link. Moving on to engagement metrics, it includes the number of followers, following count, post count, likes, and comments. These metrics serve as indicators of the account's reach, influence, and interaction levels within the Instagram community, reflecting its popularity and engagement with its audience. Additionally , metadata gives insights about timestamps of posts, post date, etc. Through careful analysis and interpretation of this data, we aim to identify factors contributing to the spread of miscellaneous activities on the platform.

B. Fake Account Detection Algorithm

The "rate user" function implements the algorithm for determining the Probability of an Instagram account being fake.It assigns a rating based on various factors such as profile picture, post count, follower-to-following ratio, posting frequency, and Activity Pattern, likes and comments for a posts. Here are the steps involved:

```
Initialize rating to 0
Initialize empty list suspicious_activities
if profile_picture is 'default' then
    Increment rating by 25
    Append "Default profile picture detected" to suspicious_activities
end if
if post_count is 0 then
    Increment rating by 25
    Append "Zero post count detected" to suspicious_activities
else if post_count is less than or equal to 5 then
    Increment rating by 10
    Append "Low post count detected" to suspicious_activities
end if
if followers_count is not None and following_count is not None then
    Calculate ratio as following_count divided by followers_count
    if ratio is greater than or equal to 2 then
        Increment rating by 30
        Append "Unusually high following count relative to followers" to suspicious_activities
    end if
end if
for each pair of consecutive post timestamps in post_timestamps do
    Calculate the time difference between the pair of timestamps
    if the time difference is less than 10 minutes then
        Increment rating by 5
        Append "Multiple posts within a short time frame" to suspicious_activities
    end if
end for
if rating is greater than or equal to 100 then
    Set rating to 100
end if
return rating and suspicious_activities
```

C. Fraud Score Calculation & Comment variable

comment variable

The variable "comments_variable" is used to store comments or remarks about the Instagram account being analyzed

Table 1
Parameter's Weights

Sr. No.	Parameters	Weightage	
		if Present	if Not Present
1	Profile Picture	0	25
2	Activity Pattern(multiple post within a day)	5	0
3	Followers vs Following	0	10 (following-followers>= 100)
		0	30(Following>2*Followers)
4	Number Of Post	10(1<= post count<=5)	25

and is initialized with "User Detail." These comments provide additional context or insights into why certain characteristics of the account may indicate it is potentially fake or suspicious. "comments_ variable" provides a summary of the analysis performed on the Instagram account, highlighting potential red flags or suspicious behaviors observed during the analysis. This summary can be useful for users or administrators to understand why a particular rating was assigned to the account.

In order to determine the fraud score, we need to assign weight to the respective parameters (profile picture, followers vs. following, number . of posts, activity pattern). In the above table, we set certain criteria and assign scores to each and every parameter. Besides, it also generates the comment on the given account and is named "comment variable." Here will be the comments and weights for the parameters:

If the profile picture of the Instagram account is detected to be a default image, this indicates a higher probability of it being fake. It will be commented as "Default image detected. Suspicious account." If the post count of the Instagram account is zero, it suggests that the account may be inactive or unused. Comment as "Zero post count. Potentially inactive account". If the post count of the Instagram account falls between 1 and 5, it suggests that the account is not very active. Comment as "Post count less than 5. Less active user." If the difference between the number of followers and followers is greater than or equal to 100, it suggests an imbalance in the follower-to-following ratio. This could indicate that the account is engaged in spammy behavior such as followforfollow schemes, indicating a higher probability of it being fake. If it's observed that following_count is double that of followers_count, comment as "unusually high following count. Potential spam or fake account". Otherwise, comment, "Unusually high following count. Potential spam or fake account" and will print (following_count vs. followers_count). If there are multiple posts within the same day with a time difference of less than 10 minutes between them, it indicates potential automated activity. comment, "Multiple posts within 10 minutes on count date times list. Potential automated activity." [Count_date_times_list represents the list of datetime objects for which multiple posts were detected within the specified time frame.]

D. Data Storage and Display

The application that we implement stores user data, including analyzed Instagram accounts and their ratings (which give detailed account insights), in a database. It also provides "Home" and "search_user" for users to interact with the system, search for accounts, and view analysis results. The application includes error handling for various scenarios, such as failed login attempts or missing data.

IV. RESULT

In this research, we present a novel approach to user authentication and fraud detection on Instagram, aimed at enhancing user trust and safeguarding against fraudulent activities.

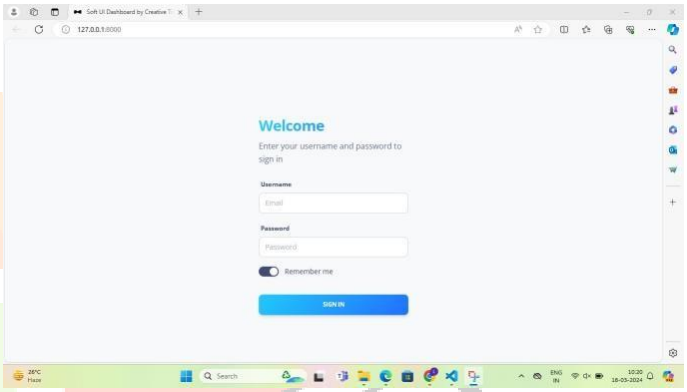

Fig. 3. Admin logging .

In figure 4, we provided an admin login page for user authentication. On this page, users have to sign in with their Gmail credentials, and login information will be stored, ensuring that only authorized users can access the system. In figure 6, upon submitting the login form with valid credentials, the user is authenticated and redirected to the dashboard.

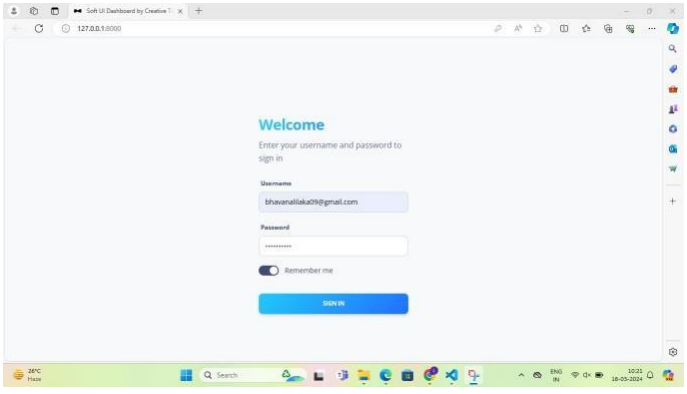

Fig. 4. Admin Logging after putting credentials.

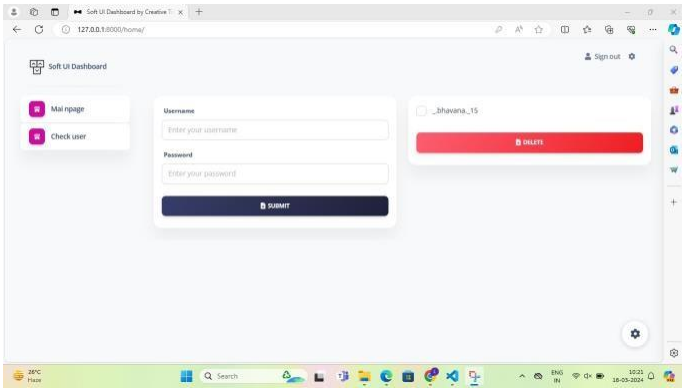

Fig. 5. Dashboard.

Dashboard provides "main page" and "check user" to the user. On the main page, the user has to login to their Instagram account. The user can manage their logging session and perform actions such as log out. The logins view employs Selenium to interact with the Instagram website and retrieve relevant data about the specified user. Submitting login details, the user is supposed to go to the check user section. This system provides "search user" for the user, which allows users to search specific Instagram accounts after putting just their username in it.

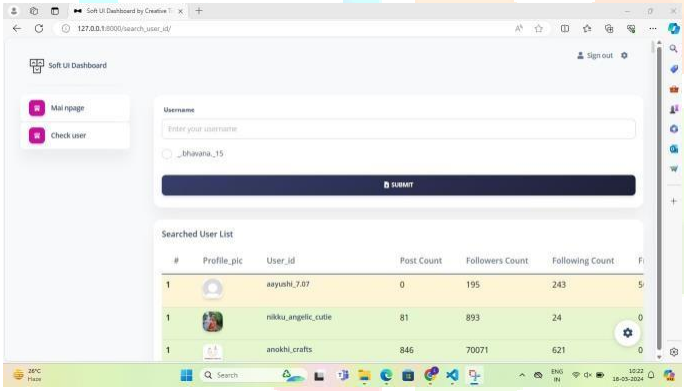

Fig. 6. In check user, Insta login page.

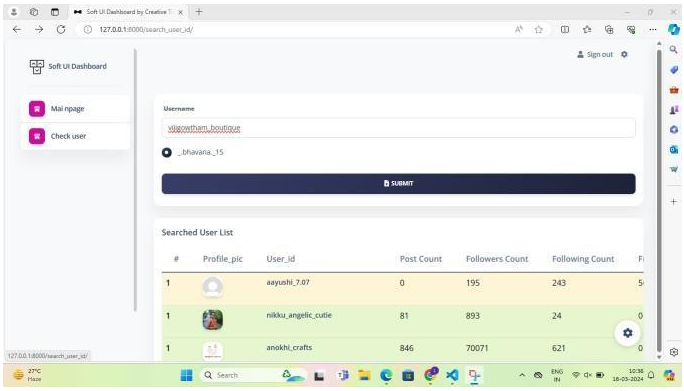

Fig. 7. Putting an account to be checked.

In figure 7, upon submission, the system retrieves information about the specified user's profile, including post count, followers, and following count. Additionally, the system analyzes the retrieved data to assign a rating to the user, indicating the fraud score of the account as suspicious or exhibiting automated activity. User-specific data, including analyzed Instagram profiles and their ratings, is stored for future reference and analysis.

For accuracy calculation, we used following notations:

Table 2
Notation

Predicted Result	Actual Result		
		Positive	Negative
	True	TP	TN
	False	FP	FN

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad *100$$

TP: The real account is correctly identified as a real account.

TN: For the real account, the system generates the warning or identifiers as a fake account.

FP: A fake account is identified as fake.

FN: fake account identified as real or generate the warning

In comparison to previous studies, our approach introduces a method using Selenium for real-time data collection and analysis directly from the Instagram platform. By analyzing parameters, achieved an accuracy rate of 91% across 220 profiles.

V. Conclusion Future Work

Fake accounts are rapidly created on social media platforms, to create spam and fraud. To fix this problem, various machine learning algorithms are used in previous research. In our current work, we tried a different approach, called selenium automation. In our project, we successfully integrated Gmail login for enhanced authentication. Then, we used a tool called Selenium to automate some tasks and collect information from websites. We also built a system that can quickly spot fake profiles online. It looks at things like profile pictures, how active someone is, how many followers they have compared to how many people they're following, and then decides if a profile is likely fake or real. And we achieve a remarkable 91% accuracy in fraud detection. Our project makes the internet safer by stopping fake accounts from causing problems. It helps make the internet safer by stopping fake accounts. It's not perfect, especially on big websites, but even with challenges, it's making a big difference in keeping the internet safe and trustworthy.

In the future, we aim to enhance our project by leveraging AI and machine learning to surpass the current accuracy rate of

91% in fraud detection. Additionally, we will devise improved methods for data collection, reducing reliance on Selenium and web scraping. Our approach to verifying account authenticity will evolve, incorporating a broader range of online behavioral indicators. Furthermore, we will extend our system's capabilities to encompass various social media platforms. Continuous user feedback will drive ongoing improvements, ensuring compliance with data protection regulations. Moreover, we will empower users with knowledge to identify online fraud. These enhancements will render our project more effective, adaptable, and dedicated to fostering a safer online environment for all.

REFERENCES

[1] Fatih Cagatay Akyon, M. Esat Kalfaoglu, "Instagram Fake and Auto- mated Account Detection", IEEE Xplore, 2019.

[2] Ananya Dey, Hamsashree Reddy, Manjistha Dey and Niharika Sinha, "DETECTION OF FAKE ACCOUNTS IN INSTAGRAM USING MACHINE LEARNING", International Journal of Computer Science Information Technology (IJCSIT), 2019.

[3] S. P. Maniraj, Harie Krishnan G, Surya T, Pranav R , "Fake Account Detection using Machine Learning and Data Science", International Journal Of Innovative Technology and Exploring Engineering (IJITEE), 2019.

[4] Preethi Harris, J Gojal, R Chitra, S Anithra, "Fake Instagram Profile Identification and Classification using Machine Learning", IEEE Xplore, 2021.

[5] Krutika Palav, Pranali Awari, Siddhi Jiman, "INSTAGRAM FAKE ACCOUNT DETECTION", International Research Journal of Modern- ization In Engineering Technology and Science, 2021.

[6] Mr.E.Ramesh Reddy, Mrs.P.Madhavi, "DETECTING FAKE AC- COUNTS ON SOCIAL MEDIA INSTAGRAM", International Journal of Research in Engineering, IT and Social Sciences, 2022.

[7] Dr. Suchita Amey Bhovar, "A Study of Different Methodologies to Detect Fake Account on Social Media using Machine Learning", In- ternational Journal of Science and Research (IJSR), 2022.

[8] Shamim Ahmad and Dr. Manish MadhavaTripathi, "A Review Article on Detection of Fake Profile on Social-Media," International Journal of Innovative Research in Computer Science Technology (IJIRCST), vol. 11, no. 2, pp. 44-49, March 2023.

[9] P. Chakraborty, M. M. Shazan, M. Nahid, Md. K. Ahmed, and P. C. Talukder, "Fake Profile Detection Using Machine Learning Techniques," Journal of Computer and Communications, vol. 10, pp. 74-87, Oct. 27, 2022.

[10] G. Jadhav, K. Patel, and Mr. R. Gawande, "Detecting Fake Accounts on Social Media using Neural Network," International Journal of Creative Research Thoughts (IJCRT), vol. 9, no. 11, pp. c56, November 2021.

[11] R. Suganya et al., "Detect fake identities using improved Machine Learning Algorithm," in Journal of Physics: Conference Series, vol. 1916, p. 012056, 2021.

[12] A. Romanov, A. Semenov, O. Mazhelis and J. Veijalainen, "Detection of Fake Profiles in Social Media – Literature Review," in Proceedings of the 13 th International Conference on Web Information Systems and Technologies (WEBIST 2017), pp. 363-369, ISBN: 978-989-758-246-2, SCITEPRESS – Science and Technology Publications, Lda, 2017, doi: 10.5220/0006362103630369.

[13] M. Mohammadrezaei, M. E. Shiri, and A. M. Rahmani, "Identifying Fake Accounts on Social Networks Based on Graph Analysis and Classi- fication Algorithms," Computational Intelligence and Neuroscience, vol. 2018.

[14] S. K. Uppada, K. Manasa, B. Vidhathri, R. Harini, and B. Sivaselvan, "Novel approaches to fake news and fake account detection in OSNs: user social engagement and visual content centric model," Social Net- work Analysis and Mining, vol. 12, no. 52, pp. 1-18, May 2022.

[15] G. Jadhav, K. Patel, and R. Gawande, "Detecting Fake Accounts on Social Media using Neural Network," International Journal of Creative Research Thoughts (IJCRT), vol. 9, no. 11, pp. c56-c62, November 2021. ISSN: 2320-2882.

[16] A. S. Chamria, A. D. Mane, S. Bharne, and P. V. Dambal, "Detecting Fake Profile in Online Social Networks using EnsemStack Classification Algorithm," in Proceedings of the 2022 6 th International Conference On Computing, Communication, Control And Automation (ICCUBEA), Pimpri Chinchwad College of Engineering (PCCOE), Pune, India, Aug 26-27, 2022, pp. 1-6. DOI:10.1109/ICCUBEA54992.2022.10010723.

[17] N. C. Leˆ, M. T. Dao, H. L. Nguyn, T. N. Nguyn, and H. Vu˜, "An Appli- cation of Random Walk on Fake Account Detection Problem: A Hybrid Approach," IEEE Transactions on Network Science and Engineering, vol. 9, no. 1, pp. 1-12, Jan. 2022. DOI: 10.1109/TNSE.2022.10010723.

[18] H. N. Khan and R. Singh, "Fake Profile Identification in Online Social Networks Using Machine Learning," Project Report, SRM Institute of Science and Technology, Kattankulathur, India, May 2021.