International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 06 | June - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

FAKE PROFILE DETECTION ON SOCIAL NETWORKING WEBSITES USING MACHINE LEARNING

Imran Mir

Assistant Professor, Guru Nanak Institute of Technology, CSE Department, Hyderabad

ABSTRACT:

In an age where social media has become an integral part of our lives, the challenge of detecting fake accounts on platforms like Instagram has gained significant importance. This project, titled "Instagram Fake Account Detection using Machine Learning," employs Python as its primary tool to tackle this problem. It leverages two powerful machine learning algorithms, the Random Forest Classifier and the Decision Tree Classifier, to accomplish this task. The Random Forest Classifier demonstrates remarkable performance, achieving a 100% accuracy on the training dataset and an impressive 93% accuracy on the test dataset. Meanwhile, the Decision Tree Classifier exhibits its effectiveness with a training accuracy of 92% and a test accuracy of 92%. The dataset employed in this project is composed of 576 records, each characterized by 12 distinct features. These features encompass critical aspects of Instagram profiles, including the presence of a profile picture, the ratio of numerical characters in usernames, the breakdown of full names into word tokens, the ratio of numerical characters in full names, the equality between usernames and full names, the length of user bios, the existence of external URLs, the privacy status of accounts, the number of posts, the count of followers, the number of accounts followed, and the ultimate classification of an account as "Fake" or "Not." By harnessing the capabilities of Python and these advanced machine learning models, this project endeavors to provide a robust and efficient solution for the identification of fake Instagram accounts. In doing so, it contributes to the preservation of the platform's integrity and the security of its users.

I. INTRODUCTION:

A website known as a "social networking site" is one where users may connect with friends, make updates, and find new people who have similar interests. Each user has a profile on the website. Users can communicate with one another using Web 2.0 technologies in these online social networks. The utilisation of social networking sites is expanding quickly and affecting how individuals interact with one another. Online communities bring together people with like interests and make it easy for users to find new friends. The main benefit of internet social networking is that it allows user to easily connect with people and communicate better. This has provided new avenues for potential attacks such as fake identities, disinformation, and more. Researchers are working to determine the impact these online social networks have on people. There is much more to media than just how many people use it. This suggests that the number of fake accounts has grown throughout the past years. The objective of this research is to develop a machine learning-based solution for detecting fake profiles on social networking websites. This entails creating algorithms that can analyze various aspects of user profiles, including attributes, behavior patterns, and engagement metrics, to accurately identify fraudulent accounts. Techniques such as natural language processing will be employed to analyze profile descriptions and textual content, while anomaly detection algorithms will detect irregular activity indicative of fake profiles. Additionally, deep learning models may be utilized to extract features from profile images for further analysis. The aim is to design a scalable and efficient system capable of real- time detection and removal

of fake profiles, collaborating with social networking platforms to integrate the solution into their existing security infrastructure. Evaluation will be conducted using relevant metrics, and user feedback mechanisms will be implemented for continuous improvement and adaptation to evolving threats.

II. LITERATURE SURVEY:

Karunakar et al. [1] draw attention to the widespread problem of online impersonation and phony accounts on social media platforms, citing a Facebook report that stated 583 million phony accounts were removed in the first quarter of 2018 alone, with an estimated 3-4% of active accounts remaining fraudulent. The goal of the suggested study is to overcome this difficulty by presenting a model that can differentiate between phony and real accounts. The model reduces the requirement for human account review by processing big datasets in an efficient manner using Support Vector Machine (SVM) as a classification approach. The project's main goal is to detect and categorize phony accounts; it approaches this challenge as a classification or clustering problem. In contrast to emails, short messaging service (SMS) is becoming more and more common in communication, as noted by Roy et al. [2]. Nevertheless, people are finding SMS spam to be an annoyance, which has prompted research into screening methods. This article uses deep learning techniques, notably Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) models, to categorize spam and non-spam communications, in contrast to prior studies that relied on manually determined criteria. These algorithms extract features on their own and only need text data. Utilizing a dataset consisting of 4,827 non-spam and 747 spam messages, the suggested method attains a remarkable 99.44% accuracy rate. The frequency of online romance frauds, a notable kind of mass-marketing fraud that is especially common in Western areas, is discussed by Tangil et al. [3]. Though common, not many research have provided data-driven remedies to address this problem. In online romance scams, con artists create accounts and manually interact with victims. Standard detection techniques, such as spam filtering, are useless because of the special features of this scam and the way dating services function. This study looks at the demographics, profile descriptions, and photo characteristics of fraudulent online dating profiles. By doing so, it sheds insight on the tactics used by scammers to lure victims as well as the characteristics of the victims themselves. A technique to automatically identify romance fraudsters on online dating sites is developed by the authors in response to the serious financial and psychological harm that dating fraud causes. The goal of the article is to create an early detection system to stop romance fraudsters from building phony profiles or corresponding with prospective victims. It provides the first completely documented system for this purpose. According to earlier studies, those who fall prey to romance frauds often score well on tests of idealized romantic ideals. The authors create a detection method to solve this by combining structured, unstructured, and deep-learned data. In a hold-out validation set, their ensemble machine-learning strategy achieves high accuracy (97%) and shows resilience to the deletion of profile information. This approach helps to reduce online dating fraud by making it easier to design automated solutions for dating site providers and individual users. The emergence of Online Social Networks (OSNs) as venues for social interaction and product/service recommendations is covered by Khan et al. [4]. On the other hand, as OSN use has increased, so too have undesired behaviors like unsolicited blogging and spamming, which may jeopardize recommendation systems and endanger authorized users. The authors provide a framework to distinguish between real experts in a certain field on Twitter and spammers and unwanted bloggers in order to solve this problem. Their method looks at twitter features and domain-specific keywords to identify unsolicited bloggers based on tweet content using a modified version of Hyperlink Induced Topic Search (HITS). The outcomes of the experiments show that the suggested technique is beneficial when compared to other cutting-edge methods and classifiers.

The relationship between users' psychological attributes, such as personality qualities, and cybercrimes such cyberbullying is examined by Balakrishnan et al. [5]. This research aims to create an automated method for detecting

cyberbullying by using the psychological characteristics of Twitter users, such as their personalities, sentiments, and emotions. The Big Five and Dark Triad models are used to evaluate the personalities of users, while machine learning classifiers like Naïve Bayes, Random Forest, and J48 are used to classify tweets into bully, aggressor, spammer, and normal categories. 5453 tweets were gathered using the #Gamergate hashtag and painstakingly annotated by human specialists to create the Twitter dataset. Certain Twitter-based characteristics, including as text, user, and network-based information, are used by baseline algorithms. Results show that adding feelings and personalities improves cyberbullying detection, but emotion does not have a comparable effect. Additional examination reveals that extraversion, agreeableness, neuroticism, and psychopathy are important characteristics affecting the identification of online bullying. Key characteristics are found using dimension reduction methods and combined into a single model to provide the best detection accuracy. The report offers recommendations for using these results to successfully reduce cyberbullying in its conclusion.

III. METHODOLOGIES:

Detecting fake profiles on social networking websites is crucial for maintaining user trust and platform integrity. Leveraging machine learning techniques, this project aims to develop an automated system capable of identifying fraudulent accounts. By analyzing various attributes such as profile information, activity patterns, and engagement metrics, the system will discern anomalies indicative of fake profiles. Natural language processing algorithms will scrutinize textual content within profiles for inconsistencies or suspicious patterns, while anomaly detection techniques will flag irregular user behaviors, such as excessive friend requests or unusual posting frequencies. Additionally, deep learning models may be employed to analyze profile images and detect any anomalies that suggest fake identities. The ultimate goal is to create a robust and scalable solution that can operate in real-time, swiftly detecting and removing fake profiles from social networking platforms. Collaborating with these platforms, the system will be integrated into their existing moderation processes to enhance their ability to combat fraudulent activities effectively. Continuous evaluation and refinement will ensure the system remains adaptive to evolving tactics employed by malicious actors, thereby bolstering user trust and maintaining the integrity of social networking communities.

EXISTING SYSTEM DISADVANTAGES:

Despite its excellent accuracy, the XGBoost algorithm has a few drawbacks when it comes to identifying phony Instagram profiles. One significant problem is that decision-making is opaque, which makes it hard for consumers to comprehend why certain accounts are reported. Furthermore, the algorithm has trouble handling unbalanced datasets, which might provide biased and inaccurate conclusions. The meticulous selection and creation of input characteristics, a process that takes time and professional expertise, is crucial to the algorithm's success.

Because the algorithm is difficult to update with fresh data, it is also less flexible in terms of responding to novel strategies used by malevolent users. Because of its high computing requirements and potential for prolonged processing periods, it is not appropriate for real-time applications. Additionally, even with high initial accuracy, there's a chance that the model won't work well with fresh data, which might lead to inaccurate account flagging.

The caliber of the data used to train the algorithm has a direct impact on its overall efficacy; any inadequacies in the data source may result in less than optimal performance. Moreover, the algorithm may not be able to identify

fraudulent accounts based just on their textual content because of its concentration on structured data. Additionally, the system does not examine various material kinds that may be exploited in misleading ways, such movies or photos. Last but not least, the algorithm's accuracy presents possible privacy issues as it may mistakenly mark real users as fraudulent, which would cause user unhappiness and distrust.

PROPOSED SYSTEM:

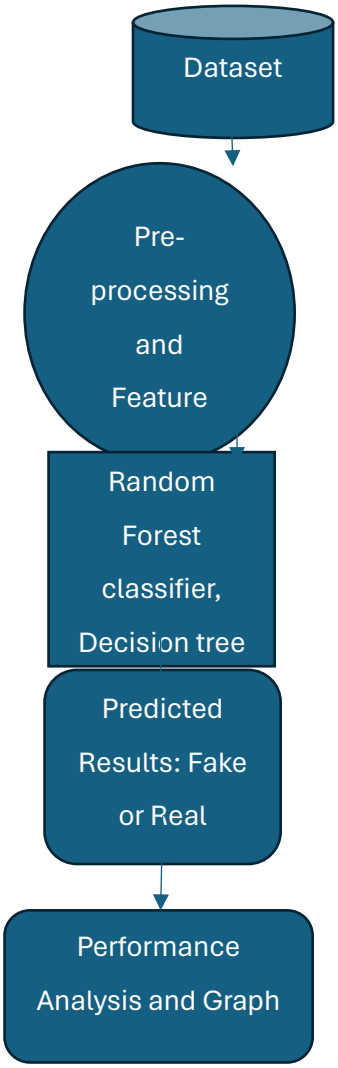
The robust programming language Python, which is well-known for its talents in data processing and machine learning, serves as the foundation for the Instagram false account identification system. The Random Forest Classifier and the Decision Tree Classifier are two well-known machine learning models that this system uses to enhance the distinction between real and phony Instagram accounts. Python is the ideal option for constructing this system because of its extensive ecosystem, which provides a multitude of modules and tools for data preparation, modeling, and assessment. These two machine learning algorithms are used by the detection system to examine Instagram accounts and establish their legitimacy. It works using a dataset of 576 records, each of which is defined by 12 unique traits that correspond to different aspects of Instagram accounts. The profile image, the number and length of the username, the words that make up the complete name, the name in the username, the description length, the external URL, the privacy setting, the number of posts, the number of followers, the number of follows, and the fake account status are some of these elements. This new system incorporates enhancements in several areas, building upon the capabilities of the previous system, which has already shown to be very accurate. It has a wide range of algorithms, strong feature engineering, and interpretability features that make it simpler to comprehend why an account has been marked as fraudulent. Additionally, the system is made to adjust to fresh dangers and malevolent actors' methods. Moreover, it prioritizes efficiency, guaranteeing its ability to handle substantial amounts of data promptly and precisely. The suggested method offers a comprehensive and efficient way to identify phony accounts, with the overall goal of improving the security and reliability of the Instagram network. By using sophisticated machine learning algorithms and inventive Python programming, it overcomes earlier shortcomings and establishes a new benchmark for the dependability of false account identification.

ADVANTAGES OF PROPOSED SYSTEM:

High accuracy is shown by the suggested method for identifying phony Instagram accounts, which makes use of both Random Forest and Decision Tree classifiers. While the Decision Tree obtains 92% accuracy on both training and test data, the Random Forest achieves 100% accuracy on both. This precision guarantees trustworthy detection of fraudulent accounts. By combining the advantages of both classifiers, the system is better equipped to handle a wide range of profile attributes. Sophisticated feature engineering methods retrieve pertinent data from Instagram accounts, offering a comprehensive perspective of user activity and enhancing the precision of detection. Included are techniques for model interpretability and explainability, which increase system confidence by enabling administrators and users to comprehend the reasons behind accounts being marked as fraudulent. Additionally, the system is built to evolve in response to new threats, with data monitoring and frequent model retraining to keep abreast of hostile actors' most recent strategies. In order to make the system scalable and effective and able to handle high volumes of profiles in real-time or near-real-time settings, efforts have been made to optimize computing resources and shorten inference times. In order to identify accounts that use textual content or multimedia manipulation, the system combines text and picture analysis, offering a thorough evaluation of account authenticity. Prioritizing user experience and privacy, safeguards are in place to reduce the possibility of incorrectly flagging authentic accounts, preserving user happiness and confidence. Data balancing techniques ensure that the system does not favor one class over another by addressing the issues associated with uneven datasets. Cross-validation and cross-referencing of

findings are made possible by the employment of both Random Forest and Decision Tree classifiers, which promotes more certain and precise detection. Strong generalization capabilities are shown by the system's high accuracy on test data, which lowers the likelihood of overfitting. As a result, there is a lower chance of false positives and negatives, establishing the system as a complete and reliable method of identifying phony Instagram accounts and enhancing the safety, reliability, and confidence of the platform.

System Architecture:



IV. Implementation:

The main interface, login interface, dataset upload, preview page, and prediction phase are the five key components that make up our false profile detection system. The primary interface functions as the user's point of entry, offering access to crucial features such dataset upload, preview, prediction, and login. By securely hashing login credentials, the login interface guarantees user authentication. For further security, it incorporates multi-factor authentication and encryption. After completing the authentication process, users are able to submit datasets in different formats to the dataset upload page, where preliminary validation checks are made to ensure data integrity. After uploading the dataset, users are sent to a preview page where they may check the data and have the opportunity to make any required edits or revisions before continuing. The prediction phase, which offers options between pre-trained and user-

uploaded models and presents findings in an easy-to-understand manner with confidence ratings for each prediction, allows users to run the machine learning model to identify phony profiles.

V. Experimental Results:

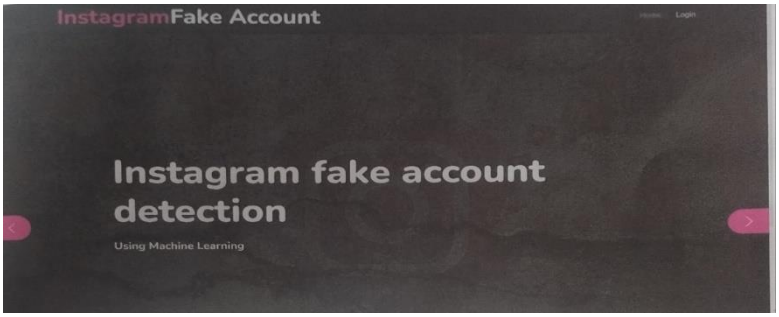


Figure. Main Interface

Figure. Login Interface

Figure: Uploading Dataset

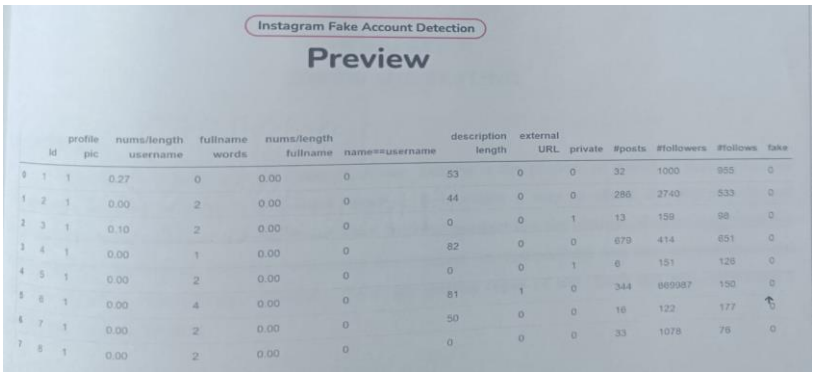


Figure: Preview Page

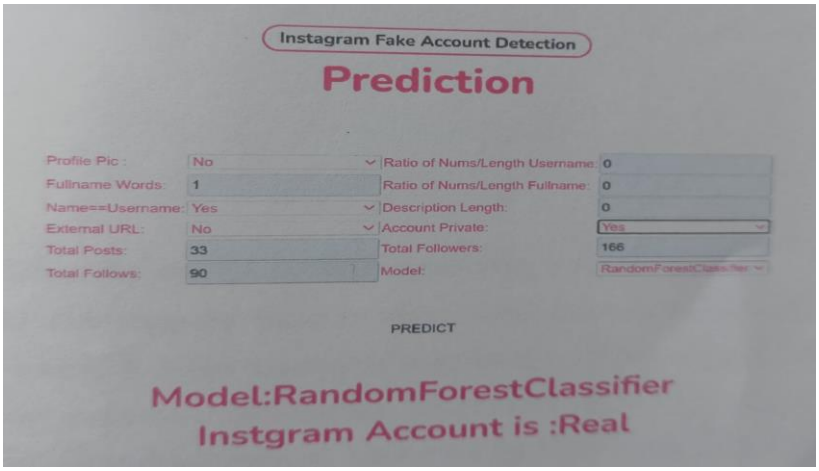


Figure: Prediction

The proposed system harnesses the power of two machine learning algorithms, the Random Forest Classifier and the Decision Tree Classifier, to collectively evaluate Instagram profiles for authenticity. Random Forest Classifier model achieves a remarkable 100% accuracy on the training dataset and a strong 93% accuracy on the test dataset, demonstrating its ability to generalize well and make accurate predictions. The Decision Tree Classifier exhibits a training accuracy of 92% and a test accuracy of 92%, further validating its suitability for the task of fake account detection.

VI. Conclusion:

To sum up, the project "Instagram Fake Account Detection using Machine Learning" offers a thorough and practical way to deal with the problem of telling real Instagram accounts from fraudulent ones. This system, which was created using Python and makes use of the Random Forest Classifier and the Decision Tree Classifier, two potent machine learning models, has shown to be very accurate and dependable in its operation. The 576 entries in the dataset that the system uses to work are enhanced with 12 unique attributes that represent different elements of Instagram accounts, including the existence of profile images, the format of usernames and complete names, the length of the bio, external URLs, and more. The system can identify phony accounts with accuracy and consistency because to these capabilities and strong feature engineering. Improvements in content analysis, interpretability, adaptation to new threats, and privacy concerns all add to the system's effectiveness and user confidence. Using a variety of

classifiers and ensuring algorithm diversity allows for a more thorough assessment of Instagram accounts. The Random Forest Classifier's test accuracy is at 93%, while the Decision Tree Classifier's test accuracy is at 92%. These results demonstrate the proposed system's great capacity to generalize and reduce false negatives and positives. These characteristics are necessary to keep the Instagram platform safe and secure. As a result, the project solves the shortcomings of the current system in addition to enhancing its strengths. It provides a comprehensive solution with the goal of improving algorithms.

VII. Future Enhancement:

The "Instagram Fake Account Detection using Machine Learning" project has established a solid framework for Instagram fake account detection, however there are still a number of areas that might need more development and enhancement to maximize the system's potential. The system should undergo regular updates and improvements as part of its continuous maintenance. These improvements should include feature engineering enhancements, model upgrades, and model generalization to better accommodate evolving threats and user behaviors. Advanced behavioral analysis may provide deeper insights into user authenticity and aid in the detection of sophisticated fake accounts. Examples of this kind of analysis include sentiment analysis and temporal study of posting habits. The accuracy of the system may be increased by implementing a method for users to report questionable accounts and by integrating user input, since user-reported data might be useful for discovering new patterns of fraudulent account activity. Its efficacy may be further increased by extending its analysis to multi-media information, including photos and videos, looking for indications of manipulation, deepfakes, and other dishonest tactics. The authenticity of accounts can be ascertained by analyzing the relationships between them, including follower networks and interaction patterns. Additionally, maintaining platform security requires the development of real-time or near-real-time monitoring capabilities to react quickly to emerging threats and suspicious activity. In order to address issues and maintain high accuracy, it is possible to investigate ways for fake account detection that respect user privacy. Additionally, by linking the system with Instagram's official API, extra user data and activity history may be accessed, which can enhance the accuracy of false account identification. To make sure the system stays competitive, benchmarking against fresh datasets and evaluating its output against cutting-edge methods should be done on a regular basis. Scalability and distributed processing will be key components of any system adaptation to manage Instagram's expanding user base and daily data generation. The system's efforts can be bolstered by creating educational materials and awareness campaigns that tell users about the risks associated with fake accounts and how to spot and report them. Additionally, working with Instagram's security and data science teams can provide important insights, access to proprietary data, and a more comprehensive approach to solving the problem. In the rapidly changing world of artificial intelligence and machine learning, it is crucial to ensure adherence to ethical standards and privacy laws, as well as to address any possible biases in the model. These avenues for future work are essential for maintaining the system's effectiveness and relevance in the dynamic environment of social media, where fake accounts continue to adapt and evolve.

REFERENCES:

[1] E. Karuniakar. V. D. R. Pavani, T. N. 1. Priya, M. V. Sri, and K. Tinivalluru, “Ensemble fake profile Jetection using machine learning (ML)," J. Inf. Comput. Sci., vol. 10, pp. 1071-1077, 2020.

[2] P. Wanda and H. J. Jie, "Deep profile: utilising dynamic search to identify phoney profiles in online ocial networks CNN" J. Inf. Secur. Appl., vol. 52, pp. 1-13, 2020.

[3] P. K. Roy, J. P. Singh, and S. Banerjee, "Deep learning to filter SMS spam," Future Gener. Comput Syst., vol. 102, pp. 524-533, 2020.

[4] R. Kaur, S. Singh, and H. Kumar, "A modern overview of several countermeasures for the rise of spam and compromised accounts in online social networks," J. Netw. Comput. Appl., vol. 112, pp. 53- 88, 2018,

[5] G. Suarez-Tangil, M. Edwards, C. Peersman, G. Stringhini, A. Rashid, and M. Whitty, "Automatically dismantling online dating fraud," IEEE Trans. Inf. Forensics Secur., vol. 15, pp. 1128- 1137, 2020.

[6] M. U. S. Khan, M. Ali, A. Abbas, S. U. Khan, and A. Y. Zomaya, "Segregating spammers and unsolicited bloggers from genuine experts on twitter," IEEE Trans. Dependable Secure Comput., vol. 15, no. 4, pp. 551-560, Jul/Aug. 2018.

[7] V. Balakrishnan, S. Khan, and H. R. Arabnia, "Improving cyberbullying detection using twitter users’ psychological features and machine learning," Comput. Secur., vol. 90, 2020, Art. no. 101710.

[8] Georgios Kontaxis, L. Polakis, S. Ioannidis and E. P. Markatos, "Detecting social network profile cloning." 2011 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops), Seattle, WA, USA, 2011, pp. 295-300, doi: 10.1109/PERCOMW.2011.5766886.

[9] Mother Aldwairi, and Ali Alwahedi, "Detecting Fake News in Social Media Networks", Procedia Compute Science, Volume 141, 2018, Pages 215-22; https:/doi.org/10.1016/j.procs 2018.10.17

[10] Buket Erşahin, Özlem Aktaş, D. Kılınç and C. Akyol, "Twitter fake account detection," 2017 pernational Conference on Computer Science and Engineering (UBMK), Antalya, Turkey, 2017, pp. J-392, doi: 10.1109/UBMK.2017.8093420.

[11] Kumud Patel, Saijshree Srivastava, and Sudhanshu Agrahari, "Survey on Fake Profile Detection on Social Sites by Using Machine Learning Algorithm," 2020 8th International Conference on Reliability, lecom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 2020, 1236-1240, doi: 10.1109/ICRITO48877.2020.9197935

[12] Alexey D.Frunze and Aleksey A. Frolov, "Methods for Detecting Fake Accounts on the Social Network VK, 2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus), St. Petersburg, Moscow, Russia, 2021, pp. 342-346, doi: 10.1109/ElConRus51938.2021.9396670.

[13] M. BalaAnand, S. Sankari, R. Sowmipriya, and S. Sivaranjani, "Recognising fraudulent users on social networks through their nonverbal cues," Int. J. Technol. Eng. Syst., vol. 7, no. 2, pp. 157-161, 2015.

[14] A. M. Meligy, H. M. Ibrahim, and M. F. Torky, "Identifier checker tool for online social networks to identify false profiles," Int. J. Comput. Netw. Inf. Secur., vol. 9, no. 1, pp. 31-39, 2017.

[15] M. Fire, A. Elyashar, and Y. Elovici, "Friend or enemy? identification of fake profiles in internet social Social networks", Social Netw. Anal. Mining, vol. 4, no. 1, 2014, Art. no. 194.

[16] Egele, G. Stringhini, C. Kruegel, and G. Vigna, "IEEE Trans. Dependable Secure Comput., "For detecting compromised accounts on social networks," IEEE Trans. Dependable Secure Comput., vol. 14, no. 4, pp. 447-460, Jul/Aug. 2017.

[17] K. Chakraborty, S. Bhattacharyya, and R. Bag, "A survey of sentiment analysis from social media data," IEEE Trans. Comput. Social Syst., vol. 7, no. 2, pp. 450-464, Apr. 2020. 59

[18] S. Lee and J. Kim, "WarningBird: IEEE Trans. Dependable Secure Comput., "A near real-time detection method for suspicious URLs in twitter stream," IEEE Trans. Dependable Secure Comput., vol. 10, no. 3, pp. 183-195, May/Jun. 2013.

[19] H. Drucker, D. Wu, and V. N. Vapnik, "In order to classify spam, support vector machines," IEEE Trans. Neural Net., vol. 10, no. 5, pp. 1048-1054, Sep. 1999.

[20] C. Chen, Y. Wang, J. Zhang, Y. Xiang, W. Zhou, and G. Min, "Real-time drifted Twitter spam detection using statistical features," IEEE Trans. Inf. Forensics Secur., vol. 12, no. 4, pp. 914-925, Apr. 2016.

[21] C. Chen et al., "Streaming spam tweets detection using machine learning: performance evaluation," IEEE Trans. Comput. Social Syst., vol. 2, no. 3, pp. 65-76, Sep. 2015.