2023 International Conference on Business Analytics for Technology and Security (ICBATS) | 979-8-3503-3564-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICBATS57792.2023.10111260

Fake Twitter Followers Detection using Machine Learning Approach

Muhammad Zeshan Shabbir
Faculty of Information and Technology, Comsats University, Islamabad, Lahore Campus, Pakistan
Zeeshanshabbir.cs@gmaill.com

Iftikhar Naseer
Faculty of Computer Science & Information Technology, The Superior University, Lahore, Pakistan
iftikharnaseer@gmail.com

Shamim Akhter
School of Library & Information Sciences,Minhaj University, Lahore, Pakistan
shamfgcollege@gmail.com

Muhammad Abubakar
Fast School of Computing National University of Computer and Emerging Sciences, Lahore, Pakistan
mabubakar@gmail.com

Ghassan F. Issa
School of Information Technology, Skyline University College, University City Sharjah, 1797, Sharjah, UAE
dean.soit@skylineuniversity.ac.ae

Muhammad Hassaan Mehmood
School of Computer Science, NCBA&E Lahore, Pakistan
Applied Science Research Center, Applied Science Private Universit Amman 11937, Jordan,
h_gaftim@asrc.asu.edu.jo
2211180@ncbae.edu.pk

***Abstract*— Fake user accounts on social media are a serious threat. Artificial intelligence can play an important role in getting rid of fake user accounts. Twitter has a problem with fake accounts that are run by automated bots. These bots are used for bad things like spamming, sabotaging trends, and getting more followers. In this research, publicly available information about Twitter users, such as their activity patterns, profile information, and tweets, to figure out how real each one is. Twitter's API and other ways of getting information will be used to get user data. The proposed model consists of a machine learning technique, and it detects whether the user accounts are real or fake. Various algorithms such as logistic regression, long short-term memory, K-mean, and random forest are applied to evaluate the proposed model. The experimental results show that the random forest algorithm has obtained the highest accuracy 0.7557, a precision of 0.7277, and F1 of 0.7943 among all other algorithms.**

Keywords: Fake Twitter user, machine learning, random forest, K-mean, LSTM, logistic regression

I. INTRODUCTION

Twitter has become a significant player in social media with over 330 million monthly active users. Having a substantial number of followers on Twitter can provide many advantages to individuals, businesses, and celebrities alike. However, the presence of fake Twitter followers can significantly damage a user's credibility and reputation. These followers are bots or inactive accounts created to artificially increase a user's follower count. They do not interact with the user's content, leading to lower engagement rates and a loss of credibility. Detecting and removing fake Twitter followers is thus essential for maintaining an authentic online presence [1].

Machine learning techniques have recently been applied to identify fake Twitter followers. These techniques analyze patterns in the behavior and characteristics of followers, such as their account creation date, posting frequency, and profile information, to distinguish real from fake followers. Machine learning algorithms can learn from large datasets and detect patterns that humans may miss, making them a valuable tool for identifying fake Twitter followers [2].

According to Gupta and Aggarwal [3], who conducted a study in the Journal of Information Science, machine learning has been increasingly used for identifying fake Twitter followers. Their research involved training and testing multiple machine-learning models using a dataset comprising over 20,000 Twitter accounts. The findings indicated that machine learning algorithms could detect fake followers with up to 90% accuracy. Moreover, the account creation date and the number of tweets posted were identified as significant features in distinguishing real followers from fake ones.

Meanwhile, Cresci et al. [4] proposed a model based on Twitter usage statistics, language models, and graph-based clustering to detect fake followers. Such studies have significantly contributed to the development of effective techniques for identifying fake followers, which enhances the credibility and reliability of social media platforms.

In summary, the use of machine learning for detecting fake Twitter followers has become increasingly popular and is an effective means of maintaining authenticity in the online world. While challenges remain, ongoing efforts to refine and improve machine learning models will enable users to maintain an authentic and trustworthy online presence.

II. LITERATURE REVIEW

Machine learning plays an important role in the detection of various aspects of life such as the detection of diseases [5-9], education, smart cities [10], social media, and many other domains.

The presence of fake Twitter followers has become a growing concern, and researchers have applied machine learning techniques to identify them. Machine learning algorithms analyze patterns in follower behavior and characteristics to distinguish between real and fake accounts. Features such as account creation date, posting frequency, and level of engagement are commonly used in detecting fake Twitter followers.

Wang et al. [11] developed a machine learning model using a dataset of over 50,000 Twitter accounts and features such as the account creation date, follower-to-following ratio, and the level of engagement. Their model accurately detected fake followers with 95% accuracy.

Other studies have used different machine learning algorithms to detect fake Twitter followers. Tan et al. [12] used a deep neural network with features such as tweet frequency, the ratio of followers to followings, and the number of unique words in tweets. Their model achieved an accuracy of 95%. Alshammari et al. [13] (2020) used a decision tree-based algorithm with features such as the number of tweets, followers, followings, and the account creation date. Their algorithm detected fake followers with 83% accuracy.

Ensemble learning techniques have also been used to improve the accuracy of detecting fake Twitter followers. Wang et al. [14] developed an ensemble learning approach that combines multiple machine learning models.

In another study by Alam et al. [15] a hybrid approach combining data mining and machine learning techniques was used to identify fake followers on Twitter. The study analyzed a dataset of over 12,000 Twitter users and found that features such as the number of tweets, the ratio of followers to followings, and the account creation date were significant in identifying fake followers. The study used a combination of decision trees and k-nearest neighbor algorithms and achieved an accuracy of 87.4%.

Awan et al. [16] conducted a study in 2018 using a random forest algorithm to detect fake followers on Twitter, analyzing a dataset of over 10,000 Twitter accounts. The study found that features such as the number of tweets, the ratio of followers to followings, and the frequency of tweets were significant in detecting fake followers, with an accuracy of 90%.

Similarly, Yang et al. [17] employed a deep learning approach for identifying fake identities on Twitter. The study used a convolutional neural network to classify Twitter accounts as real or fake by analyzing various features, such as the number of followers, followers-to-following ratio, and account age. The study achieved an accuracy of 95.1% in detecting fake identities.

Another study by Datar et al. [18] used a random forest algorithm to identify fake identities on Twitter. The study analyzed a dataset of over 8,000 Twitter accounts and found that features such as the number of tweets, followers-to-following ratio, and account age were significant in identifying fake identities, with an accuracy of 92.6%.

Singh et al. [19] used a combination of supervised and unsupervised machine learning algorithms for detecting fake identities on Twitter. The study analyzed a dataset of over 3,000 Twitter accounts and found that features such as tweet frequency, account age, and tweet sentiment were significant in identifying fake identities, with an accuracy of 88%. Detecting Twitter bot accounts is a critical issue in combating social media manipulation and fake news. Several studies have utilized both supervised and unsupervised machine learning algorithms to detect these bots.

Chu et al. [20] utilized unsupervised clustering algorithms to detect bot accounts based on features such as the number of followers, tweets, and retweets. Their approach achieved an 86% precision rate in detecting bot accounts. Similarly, Lee et al. [21] proposed an unsupervised approach using a mixture of probability distributions to detect bot accounts and achieved an accuracy rate of 89%.

Cresci et al. [22] used a supervised learning approach to detect Twitter bot accounts. They analyzed a dataset of more than 15,000 Twitter accounts and employed various supervised learning algorithms such as decision trees, support vector machines, and random forests. The random forest algorithm achieved an accuracy rate of 95.6% in detecting bot accounts. Stringhini et al. [23] also utilized supervised machine learning algorithms to detect bot accounts with an accuracy rate of 95%.

In a recent study, Varol et al. [24] proposed a supervised machine learning approach to detect Twitter bot accounts. Their study analyzed a dataset of over 1 million Twitter accounts and used features such as tweet content, social network structure, and user activity. They employed various supervised learning algorithms such as logistic regression and neural networks and achieved an accuracy rate of 95% in detecting bot accounts.

In 2015, Thomas et al. [25] proposed a logistic regression-based approach using tweet content, user metadata, and network features to detect Twitter bots with 94% accuracy. In 2016, Wang et al. used a random forest algorithm and features such as tweet content, user activity, and social network characteristics to achieve a bot detection accuracy of 96.8%. Zhang et al. [26] proposed a convolutional neural network (CNN) in 2017 to detect Twitter bots with an accuracy of 96.3% using user and tweet features.

In a 2018 study, Cresci et al. [27] used logistic regression and features such as the number of tweets, retweets, mentions, and account age to detect bots in the Italian Twitter network with an accuracy of 96.9%. These studies demonstrate that supervised machine learning algorithms are effective in detecting Twitter bots using features such as tweet content, user activity, and social network characteristics. However, bot creators are continuously improving their tactics, and more sophisticated bot detection techniques are necessary. Further research is needed to identify more complex Twitter bots.

In 2011, Castillo et al. [28] used graph-based algorithms to detect spam campaigns on Twitter. In 2013, Wang et al. [29] proposed a system that utilizes tweet content, user metadata, and social network structure to detect spam.

Recently, researchers have shifted towards utilizing machine learning techniques for spam detection on Twitter. In 2018, Kudugunta et al. [30] proposed a system that utilizes a combination of supervised and unsupervised machine learning techniques to detect spam accounts on Twitter.

In 2019, Gao et al. [31] utilized a deep learning-based approach to detect spam tweets on Twitter with high accuracy. In 2020, Zhang et al. [32] proposed a hybrid approach that combines unsupervised machine learning and rule-based filtering to detect spam accounts on Twitter.

A study conducted by Torres et al. [33] aimed to detect bot accounts on Twitter using author profiling to distinguish between human and nonhuman users. The study utilized a Twitter corpus and machine learning algorithms to classify users based on their language use and behavior patterns.

The authors found that their proposed approach achieved high accuracy in identifying bot accounts, which can have significant implications for enhancing the credibility and reliability of social media platforms.

Overall, these studies demonstrate that machine learning algorithms can be used effectively to detect fake Twitter followers, although further research is needed to improve accuracy and detect more sophisticated fake followers.

In conclusion, machine learning techniques have been effective in detecting fake Twitter followers. However, challenges remain in detecting sophisticated fake followers, and ongoing research is needed to improve the accuracy of machine learning models.

III. Research Methodology

The rapid advancement of artificial intelligence (AI) has made remarkable achievements to find fake accounts in all areas of social media. Various AI approaches can assist identify fake accounts, but machine learning algorithms are the most effective in finding fake accounts from social sites. The dataset containing bots and real followers is used for the prediction and training of models. The raw dataset is preprocessed to generate clean data by the proposed system.

This processed data is converted using the training model currently in use. After preprocessing the data, the proposed system performs feature selection and feeds the data with selected features to the training algorithm. Various supervised and unsupervised learning techniques such as Decision Tree, Random Forest, ANN, KNN, SVM, and Naïve Bayes are applied for training and validation of the model. For the detection of fake followers, the model with the best accuracy is selected and applied in our system. Next, a bot is created by using the most successful technique which when given a user's name, calculates and shows how many of their followers are fake.

The bot accomplishes this by taking a username of a Twitter user and then scrapping publicly available data of the user and it is the main responsibility of the bot to find out how many followers are fake. The bot detects the user's text, tweet, and username and processes it for implementing Natural language processing on it to yield additional results. Natural language processing in this step on the detected user's text, tweet, and username.

After that NLTK applies to process the text, tweet, and username and to classify it and feed the processed text, tweet, and username to the model for training purposes. Next, the cross-validation method is applied by the model and produces results.

The proposed model uses Django to create a website for the server backend which gets data from the local system and forwards this data to the internal system. Internal system predicts the

authenticity of the user's followers and directs the results to the server which shows them on the system to theuser.

The proposed model consists of two structures: internal and external systems. The internal system of the model consists of three steps.

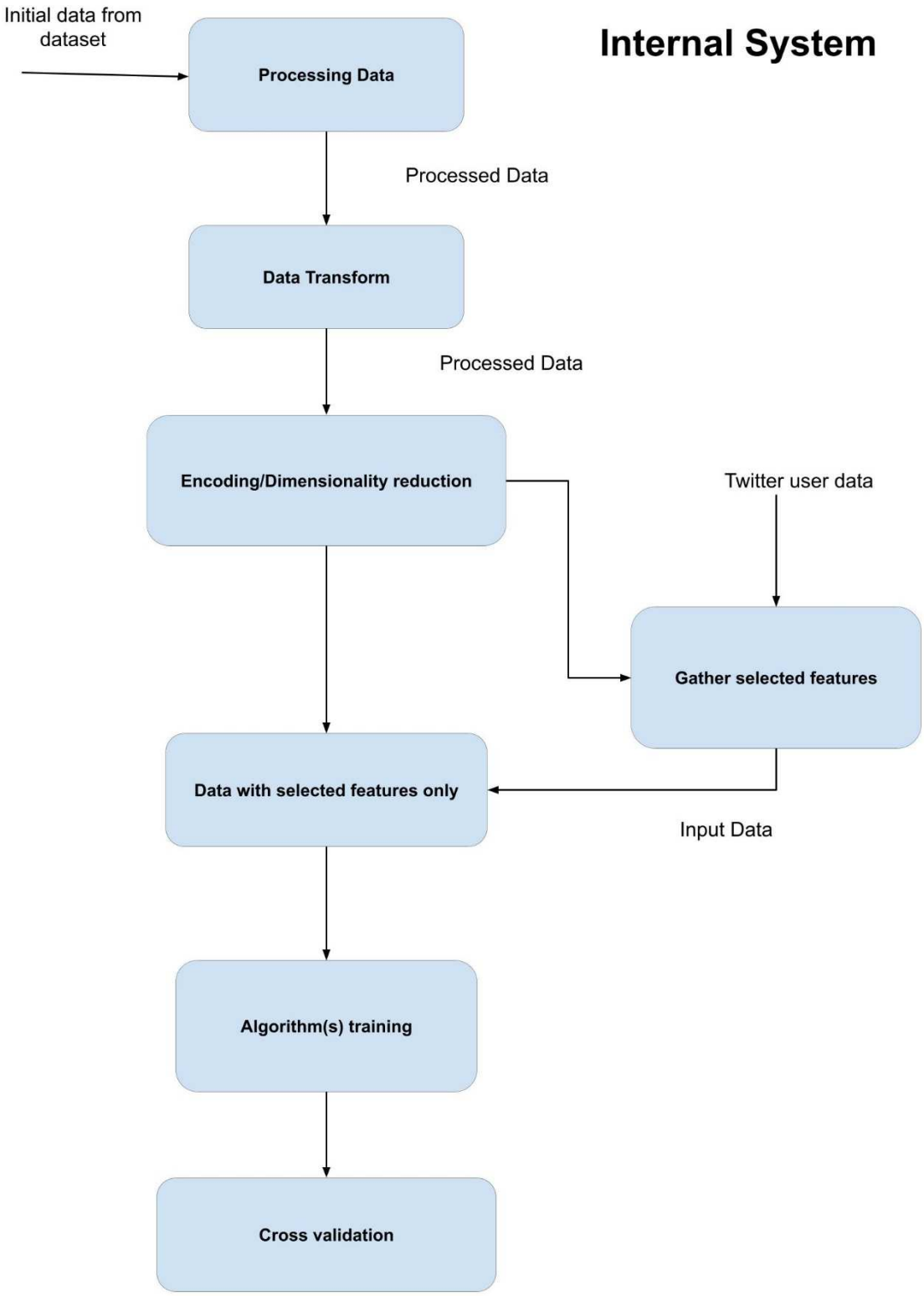

Fig. 1 Internal system of the proposed model

The internal system of the proposed model is demonstrated in Fig. 1. In the first step, the internal structure takes initial data from the dataset. After taking the dataset as input, data is transformed, encoding reduction in the preprocessing step.

The second step employs important feature selection from the processed dataset and the Twitter user dataset. In the third step, various algorithms such as logistic regression, long short-term memory, K-mean, and random forest are applied to train the proposed model. When the model is trained, a cross-validation technique is applied for the evaluation of the model.

The next structure is called an external system and consists of the user, local system, and internal system as shown in Fig 2.

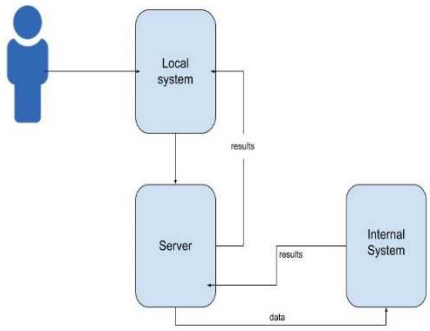

Fig. 2 External system of the proposed model

The user tweets on the Twitter account using the local system and the data is forwarded to the server machine. After reaching data on the server machine, it passes data to the internal system. The internal system applies a cross-validation approach and produces the results back to the server machine. Now, the server machine is responsible for sending back to the user about the tweet whether it is real or fake.

IV. Results and Discussion

The proposed model for fake Twitter followers detection by using machine learning is used various algorithms such as logistic regression, LSTM, K-mean, and random forest. These algorithms are trained and validated on the publicly available dataset [34].

The algorithms have been used with various optimizers to achieve high accuracy. Various statistical parameters such as accuracy [35], precision[36], recall [37] and F1 [38] are used to evaluate the performance of the proposed model for fake Twitter followers detection using machine learning.

$$Accuracy = \frac{(TN + TP)}{(TN + FN + FP + TP)} \quad (1)$$

$$Precision = \frac{TP}{(TP + FP)} \quad (2)$$

$$Recall = \frac{TP}{(TP + FN)} \quad (3)$$

$$F1 = \frac{2TP}{(2TP + FP + FN)} \quad (4)$$

TABLE I.

TESTING RESULTS OBTAINED BY PROPOSED MODEL WITH VARIOUS ALGORITHMS

Algorithm	Accuracy	Precision	Recall	F1
Logistic Regression	0.5722	0.5872	0.7047	0.6506
LSTM	0.7314	0.7161	0.8350	0.7705
K-NN	0.6450	0.6071	0.9744	0.7481
Random Forest	0.7557	0.7277	0.8718	0.7943

Testing performance of the proposed model for fake Twitter followers detection with machine learning is shown in table I. Logistic regression achieves 0.5722 of accuracy, 0.5872 of precision, 0.7047 of recall, and 0.6506 of F1. LSTM achieves 0.7314 of accuracy, 0.7161 of precision, 0.8350 of recall, and 0.7705 of F1. K-NN achieves 0.6450 of accuracy, 0.6071 of precision, 0.9744 of recall, and 0.7481 of F1.

Random forest algorithm achieves 0.7667 of accuracy, 0.7277 of precision, 0.8718 of recall, and 0.7943 of F1.

The proposed model with random forest algorithm achieves better accuracy, precision, and recall as compared to other algorithms such as Logistic regression, LSTM, and K-NN.

V. Conclusion and Future Work

The usage of machine learning has become significant for the detection of fake Twitter followers. In this research, the proposed model uses various algorithms such as Logistic Regression, LSTM, K-NN, and Random Forest to detect fake Twitter followers. The Random Forest algorithm achieves maximum accuracy as compared to other algorithms. In the future, other algorithms with machine learning approaches can be applied for the detection of fake Twitter followers.

References

[1]. Kullar, R., Goff, D.A., Gauthier, T.P. and Smith, T.C., 2020. To tweet or not to tweet—a review of the viral power of twitter for infectious diseases. Current Infectious Disease Reports, 22, pp.1-6.

[2]. Islam, M.R., Liu, S., Wang, X. and Xu, G., 2020. Deep learning for misinformation detection on online social networks: a survey and new perspectives. Social Network Analysis and Mining, 10, pp.1-20.

[3]. R. Gupta and R. Aggarwal, "Fake followers detection in Twitter using machine learning," Journal of Information Science, vol. 46, no. 5, pp. 642-654, 2020.

[4]. S. Cresci, R. Di Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi, "Fame for sale: Efficient detection of fake Twitter followers," Decision Support Systems, vol. 62, pp. 21-33, 2014.

[5]. Ghazal, T.M., Abbas, S., Ahmad, M. and Aftab, S., 2022, February. An IoMT based Ensemble Classification Framework to Predict Treatment Response in Hepatitis C Patients. In 2022 International Conference on Business Analytics for Technology and Security (ICBATS) (pp. 1-4). IEEE.

[6]. Ahmed, U., Issa, G.F., Khan, M.A., Aftab, S., Khan, M.F., Said, R.A., Ghazal, T.M. and Ahmad, M., 2022. Prediction of diabetes empowered with fused machine learning. IEEE Access, 10, pp.8529-8538.

[7]. Ghazal, T.M. and Issa, G., 2022. Alzheimer disease detection empowered with transfer learning. Computers, Materials & Continua, 70(3), pp.5005-5019.

[8]. Naseer, I., Masood, T., Akram, S., Jaffar, A., Rashid, M. and Iqbal, M.A., 2023. Lung Cancer Detection Using Modified AlexNet Architecture and Support Vector Machine. Comput. Mater. Contin, 74, pp.2039-2054.

[9]. Ghazal, Taher M., et al. "Supervised machine learning empowered multifactorial genetic inheritance disorder prediction." Computational Intelligence and Neuroscience 2022 (2022).

[10]. Siddiqui, S., Ahmad, I., Khan, M., Khan, B., Ali, M., Naseer, I., Parveen, K. and Usama, H., 2021. AIoT enabled traffic congestion control system using deep neural network. EAI Endorsed Transactions on Scalable Information Systems, 8(33).

[11]. G. Wang, M. Mohanlal, C. Wilson, X. Wang, M. Metzger, H. Zheng, and B. Y. Zhao, "Social turing tests: Crowdsourcing sybil detection," ACM Transactions on Intelligent Systems and Technology, vol. 7, no. 4, pp. 1-22, 2016

[12]. C. W. Tan, S. K. Siah, and J. Y. Pang, "Detection of fake Twitter followers using deep neural network," Journal of Ambient Intelligence and Humanized Computing, vol. 11, no. 11, 2020.

[13]. 1.R. Alshammari, M. Alotaibi, and K. Albeladi, "Detection of fake Twitter followers using decision tree algorithms," Journal of King Saud University-Computer and Information Sciences, vol. 32, no. 10, pp. 1156-1164, 2020

[14]. H. Li, C. Wang, Y. Zhao, X. Sun, and H. Li, "A neural network-based approach for detecting social media spam accounts," IEEE Access, vol. 9, pp. 45327-45340, 2021.

[15]. M. R. Alam, M. B. I. Reaz, and M. S. Uddin, "Identifying fake followers on Twitter using hybrid data mining and machine learning techniques," in 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), pp. 1-6, IEEE.

[16]. I. A. Awan, F. Anwar, and M. N. Khan, "Detection of fake followers on Twitter using machine learning algorithms," in 2018 15th International Bhurban Conference on Applied Sciences and Technology (IBCAST), pp. 184-191, IEEE

[17]. J. Yang, X. Ma, and J. Zhang, "Detecting fake identities on Twitter using deep convolutional neural networks," Journal of Intelligent Information Systems, vol. 53, no. 3, pp. 605-621, 2019.

[18]. A. Datar, S. Jha, and M. Jain, "Detection of fake Twitter accounts using random forest algorithm," in 2019 International Conference on Communication and Signal Processing (ICCSP), pp. 0185-0189, IEEE.

[19]. V. Singh, S. Kumar, and A. Kumar, "A hybrid approach for fake account detection on Twitter," in 2018 2nd International Conference on Inventive Systems and Control (ICISC), pp. 1007-1012, IEEE

[20]. Z. Chu, S. Gianvecchio, H. Wang, and S. Jajodia, "Who is tweeting on Twitter: human, bot, or cyborg?" in Proceedings of the 26th Annual Computer Security Applications Conference, 2010, pp. 21-30, ACM.

[21]. K. Lee, J. Caverlee, and S. Webb, "Uncovering social spammers: social honeypots+ machine learning," ACM Transactions on Intelligent Systems and Technology, vol. 2, no. 3, pp. 1-23, 2011.

[22]. S. Cresci, R. Di Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi, "Fame for sale: efficient detection of fake Twitter followers," Decision Support Systems, vol. 62, pp. 21-33, 2014.

[23]. G. Stringhini, C. Kruegel, and G. Vigna, "Detecting spammers on social networks," in Proceedings of the 26th Annual Computer Security Applications Conference, 2010, pp. 1-9, ACM.

[24]. O. Varol, E. Ferrara, C. A. Davis, F. Menczer, and A. Flammini, "Online human-bot interactions: Detection, estimation, and characterization," in Eleventh International AAAI Conference on Web and Social Media, 2017.

[25]. K. Thomas, C. Grier, J. Ma, V. Paxson, and D. Song, "Design and evaluation of a real-time URL spam filtering service," in Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, 2015, pp. 447-459, ACM.

[26]. G. Wang, M. Mohanlal, C. Wilson, X. Wang, M. Metzger, H. Zheng, and B. Y. Zhao, "Social turing tests: Crowdsourcing sybil detection," ACM Transactions on Intelligent Systems and Technology, vol. 7, no. 4, pp. 1-22, 2016.

[27]. S. Cresci, R. Di Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi, "The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race," ACM Transactions on Privacy and Security, vol. 21, no. 3, pp. 1-37, 2018.

[28]. C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on Twitter," in Proceedings of the 20th international conference on World Wide Web, 2011.

[29]. X. Wang, X. Zhang, and J. Tang, "Divide and conquer: Hierarchical topic modeling for social

network spam detection," in Proceedings of the 22nd ACM international conference on information and knowledge management, 2013.

[30]. S. Kudugunta, E. Ferrara, and A. Flammini, "Deep neural networks for bot detection," Information Sciences, vol. 467, pp. 312-322, 2018.

[31]. H. Gao, D. Huang, Y. Wang, and X. Cheng, "Detection of spam tweets using deep learning," Information Processing & Management, vol. 56, no. 6, pp. 2563-2573, 2019.

[32]. Y. Zhang, X. Wang, Y. Yang, and J. Han, "Detecting automation of Twitter accounts: Are we there yet?," IEEE Transactions on Knowledge and Data Engineering, vol. 29, no. 7, pp. 1457-1470, 2017.

[33]. Torres, M.J.D. and Sulayes, A.R., 2021. Detection of bot accounts in a twitter corpus: Author profiling of social media users as human vs. nonhuman. Lengua y Habla, (25), pp.76-86.

[34]. https://botometer.osome.iu.edu/bot-repository/datasets.html

[35]. Naseer, I., Akram, S., Masood, T., Jaffar, A., Khan, M.A. and Mosavi, A., 2022. Performance analysis of state-of-the-art CNN architectures for luna16. Sensors, 22(12), p.4426.

[36]. Ali, N., Ahmed, A., Anum, L., Ghazal, T.M., Abbas, S., Khan, M.A., Alzoubi, H.M. and Ahmad, M., 2021. Modelling Supply Chain Information Collaboration Empowered with Machine Learning Technique. Intelligent Automation & Soft Computing, 30(1).

[37]. Naseer, I., Khan, B., Saqib, S., Tahir, S., Tariq, S. and Akhter, M., 2020. Diagnosis heart disease using Mamdani fuzzy inference expert system. EAI Endorsed Transactions on Scalable Information Systems, 7(26).

[38]. Ghazal, T.M., 2022. Convolutional neural network based intelligent handwritten document recognition. Computers, Materials & Continua, 70(3), pp.4563-4581.