A CONCEPTUAL FRAMEWORK OF A DETECTIVE MODEL FOR SOCIAL BOT CLASSIFICATION

Emmanuel Etuh[1,2], George E. Okereke[1], Deborah U. Ebem[1], and Francis S. Bakpo[1]

[1]Department of Computer Science, Faculty of Physical Sciences, University of Nigeria Nsukka, Nigeria
[2]Department of Mathematics/Statistics/Computer Science, Kwararafa University, Wukari, Taraba State, Nigeria

ABSTRACT

Social media platform has greatly enhanced human interactive activities in the virtual community. Virtual socialization has positively influenced social bonding among social media users irrespective of one's location in the connected global village. Human user and social bot user are the two types of social media users. While human users personally operate their social media accounts, social bot users are developed software that manages a social media account for the human user called the botmaster. This botmaster in most cases are hackers with bad intention of attacking social media users through various attacking mode using social bots. The aim of this research work is to design an intelligent framework that will prevent attacks through social bots on social media network platforms.

KEYWORDS

Social media platform, human user, social bot, hackers, social security, intrusion prevention

1. INTRODUCTION

Virtual socialization has greatly enhanced social bonding irrespective of one's location in the global village. Different social media platform exists to help with different aspects of social interactions. During the past decade, social media like Twitter and Facebook emerged as a widespread tool for massive-scale and real-time communication [1]. Twitter and Facebook alone attracts over 500 million users across the world [2] which shows a rapid growth in the virtual community. Two categories of social media users identified in this virtual community are human users and social bot users. While human users personally operate their social media accounts, social bot users are developed software that manages a social media account for the human user called the botmaster. A typical example of a social bot user is a twitter bot which can be automated to write tweets, re-tweet, and like a tweet. Twitter platform does not mind the use of Twitter bot accounts as long as they do not break the Terms of Service of the platform [3]. Just as there are good human social media users and bad users called hackers, there are also good social bot user that manages the botmaster's account and bad ones as well used for attacks on the social media platform. Majority of human users of the social media platform are less knowledgeable about the functionality, security features and precautionary measures necessary to enhance safe interaction in the social cyberspace. In a more convenient way, bad users called hackers preferably employ the use of bad social bots to attack unsuspecting users. Hence, differentiating between a human user and bad social bot user becomes essential to inform a naïve user on the level of trust to be given to a social connection for virtual interaction. There is a need for a more reassuring proof of identity in the global village [4]–[7]. This proof of identity verification can

either be through what the user does on the social media platform (user activities) or account features of the user. Varying attacks have been witnessed on the social media platform, these attacks have been summarized in [8].

Many researchers have proposed different security mechanisms to curtail the activities of hackers on social media platform. Some of these proposals include: biometric authentication, hybrid system for anomaly detection in social networks [9], Network Intrusion Detection System [10], [11], [12]. On social bot detection, [13] proposed "An Evolutionary Computation Approach for Twitter Bot Detection", [14] worked on "Twitter bot detection using supervised machine learning", [15] worked on "Twitter Bot Detection using Diversity Measures", [16] proposed "Twitter Bot Detection Using Bidirectional Long Short-term Memory Neural Networks and Word Embeddings", [3] proposed "Supervised Machine Learning Bot Detection Techniques to Identify Social Twitter Bots", [1] proposed "Deep Neural Networks for Bot Detection", [2] proposed "Fake Account Detection in Twitter Based on Minimum Weighted Feature set.

Being an evolving domain that is rapidly growing, different proposal for social bot detection on social media platform is still evolving. Hackers too are not relenting in developing evading techniques to detection. Hence, there is therefore a need for intelligent intrusion detection (IIDM) model that is efficient to disarm the hackers from carrying out their cybercrime activities against SMNP by promptly notifying a typical social media human user the account category of a new found user on the platform to prevent attacks through social bot developed by hackers for malicious intention. This work seeks to enhance the social media usage by exposing hackers' use of social bots which are potential wide scale attacking tool on the social media platform.

2. RELATED LITERATURES

Due to the virtual nature of the global village occasioned by the advent of the Internet [5], human users and social bot users interact in the virtual world. This mixed interaction affects virtual socialization. In this section, a theoretical background is given on social media platform and the proposed model by some researcher on how to counter attacks on the social media platform.

2.1. Theoretical Background

Social media platforms have become an integral part of average Internet users in the virtual community today. Billions of connected devices to the Internet operate on one social media platform or the other. According to report in [17], over 500 million IoT devices were implemented globally in 2003, 12.5 billion in 2010, and 50 billion in 2020. There are about 3.5 billion people on social media with an estimated attacks that generate over $3 billion annually for cyber criminals [18]. Online social network platform like Facebook incorporate several functionalities which includes product and services advertisement, and sales that makes it relevant to almost all internet users either cooperate or private. The Covid19 pandemic has been instrumental to the geometric shift to virtual socialization. Also, the technological shift to cloud computing paradigm also has positively influenced the ubiquity of social media. This has also increased cybercriminals' activity on the platform. According to a survey by CERT, the rate of cyber-attacks has been doubling every year [10]. Online social network is faced with threatening security challenges [19]. This shift seems to have given hacker an edge to securely carryout their nefarious acts since humans are less involved. Cloud intrusion attacks are set of actions that attempt to violate the integrity, confidentiality or availability of cloud resources on cloud SMNP. The rising drop in processing and Internet accessibility cost is also increasing users' vulnerability to a wide variety of cyber threats and attacks. There are two types of users of social media platforms, they are: human users and social bot users. The human users are human beings that

directly operate their social media account through connected devices while social bot users are developed software that manages a social account for a human user. These social bot users can also be categorized into two depending on the activities carried out by them to be either good social bot or bad social bot [2]. The bad social bot are malicious software designed for misuse of a targeted social media platform. Intrusion detection is meant to detect misuse or an unauthorized use of the computer systems by internal and external elements [11]. IDS are an effective security technology, which can detect, prevent and possibly react to the attack [20], [21] opined that artificial Intelligence plays a driving role in security services like intrusion detection. Several attacks on social media platforms can best be detected by developing an intelligent intrusion detection model for social media platform [8].

2.2. Review of Related Literatures

[13] proposed "An Evolutionary Computation Approach for Twitter Bot Detection". The researcher used genetic algorithms and genetic programming to discover interpretable classification models for Twitter bot detection with competitive qualitative performance, high scalability, and good generalization capabilities. The model was able to detect twitter bots with detection accuracy of 75 per cent.

[14] proposed "Twitter bot detection using supervised machine learning". They used algorithms like Decision tree, K nearest neighbours, Logistic regression, and Naïve Bayes to calculate accuracy in classifying bots and compared it with their model classifier that used bag of bots' word model to detect Twitter bots from a given training data set. The proposed classifier is based on Bag of Words (BoW) model which is used to extract features from text in the areas of Natural Language Processing or NLP, Computer Vision and Information Retrieval (IR). The twits from a user is compared with BoW to determine if the account is a bot.

[2] proposed "Fake Account Detection in Twitter Based on Minimum Weighted Feature set". Over 22 factors for determining a fake account were mined out of which the study minimized set of the main factors that influence the detection of the fake accounts on Twitter, and then the determined factors are applied using different classification techniques.

[16] proposed "Twitter Bot Detection Using Bidirectional Long Short-term Memory Neural Networks and Word Embeddings" the model used Bidirectional Long Short-term Memory Neural Networks and Word Embeddings for Twitter bot detection. The model only relies on tweets and does not require heavy feature engineering to detect bots on Twitter.

[1] proposed a "Deep Neural Networks for Bot Detection". Their model design was based on contextual long short-term memory (LSTM) architecture that exploits both content and metadata to detect bots at the tweet level. Other contextual features are extracted from user metadata and fed as auxiliary input to LSTM deep nets processing the tweet text forming a dens layer that generate the output which classifies the accounts as either bot or not.

[9] proposed "an efficient hybrid system for anomaly detection in social networks". The model cascaded several machine learning algorithms that included decision tree, Support Vector Machine (SVM) and Naïve Bayesian classifier (NBC) for classifying normal and abnormal users on social networks. the anomaly detection engine uses SVM algorithm to classify social media network user as happy or disappointed, NBC algorithm is used based on a defined dictionary to classify social media users with social tendency. Unique features derived from users' profile and contents were extracted and used for training and testing of the model, performance evaluation conducted by experiment on the model using synthetic and real datasets from social network shows 98% accuracy.

[3] proposed "Supervised Machine Learning Bot Detection Techniques to Identify Social Twitter Bots". The model combined user profile, account activity, and text mining to predict the user account as bot or otherwise using complex machine learning algorithm which utilized a range of features like length of user names, reposting rate, temporal patterns, sentiment expression, followers-to-friends ratio, and message variability for bot detection

3. CONCEPTUAL FRAMEWORK

The proposed Intelligent Intrusion Detection Model for social bot classification will follow Machine Learning (ML) design approach. Machine learning is all about programming computers to optimize a performance criterion using past experience encoded as dataset [22]. A social media user on a particular platform can verify each social contact to detect the status of the account which should influence the extent of virtual socialization with the new user. The proposed system will utilize account-level features to identify 'who the user is' in the virtual space. If an account is detected to be a social bot, the user is notified. This will serve as a preventive mechanism that will shield the user from the designed attacks of the hacker that uses bots to attack the social media user. Otherwise, if the account is detected to be a human user, then the social media user can now virtually relate with the user. Beforehand, the social media user would have fallen into this kind of attack before devising a way of recovering from the attack, but with the proposed model, the user will escape social bot related attacks.

3.1. Activity Diagram

The user, the model, and the platform are the three entities to be considered in the proposed design. The social media user triggers the activity when they want to connect with a new user. This activity triggers the model to extract the account features of the new social media user to identify the type of user it is. A web crawler will be used to extract the account features of the new user, this feature dataset will be passed to the model for detection, if the prediction of the new user is a bot, the details will be communicated to the user in a log file and the activity stops. Else, the user can now read new twits from the new user or follow the new user or interact socially with the new user without fear of attack. The activity diagram of the model is presented in Fig 1 below.

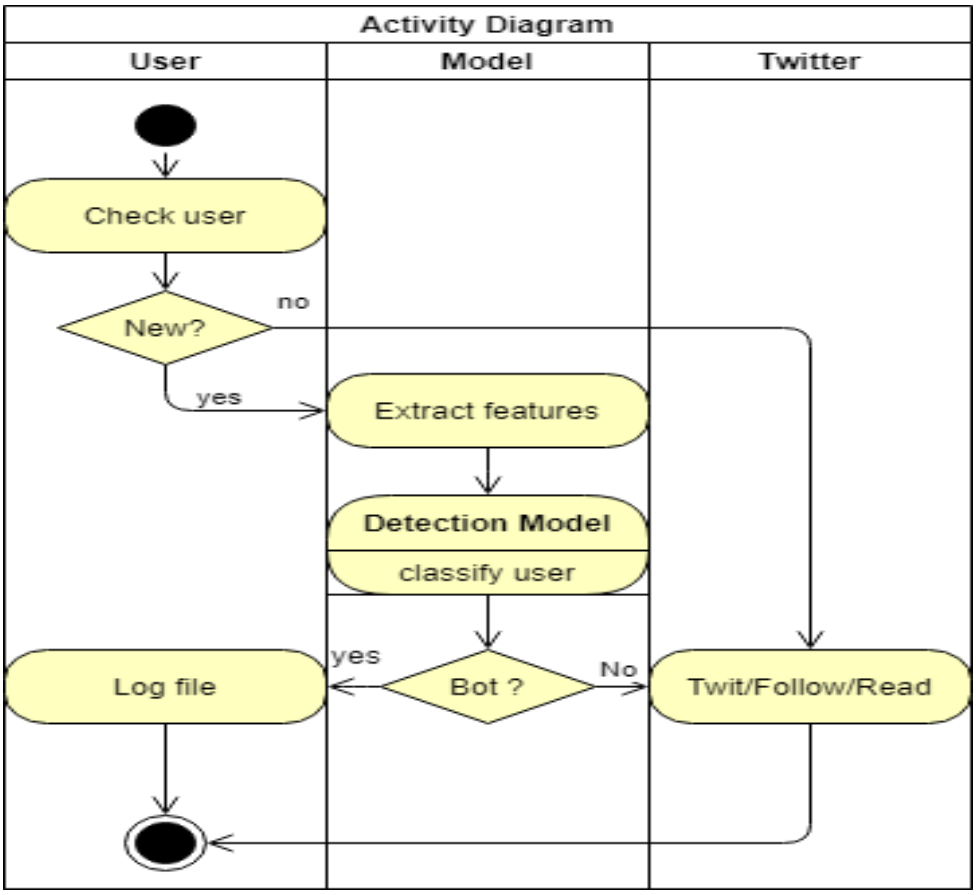


Fig 1: Activity Diagram

3.2. Flow chart

The model read the new user and extracts the user features that will be passed to the detection engine to predict if the new user is a social bot or a human being to enable the typical user make informed decision on how to relate with the user. If the model predicts the user to be a human user, the typical user can then go on to virtually relate with the user. The flowchart of the proposed model is presented in Fig 2

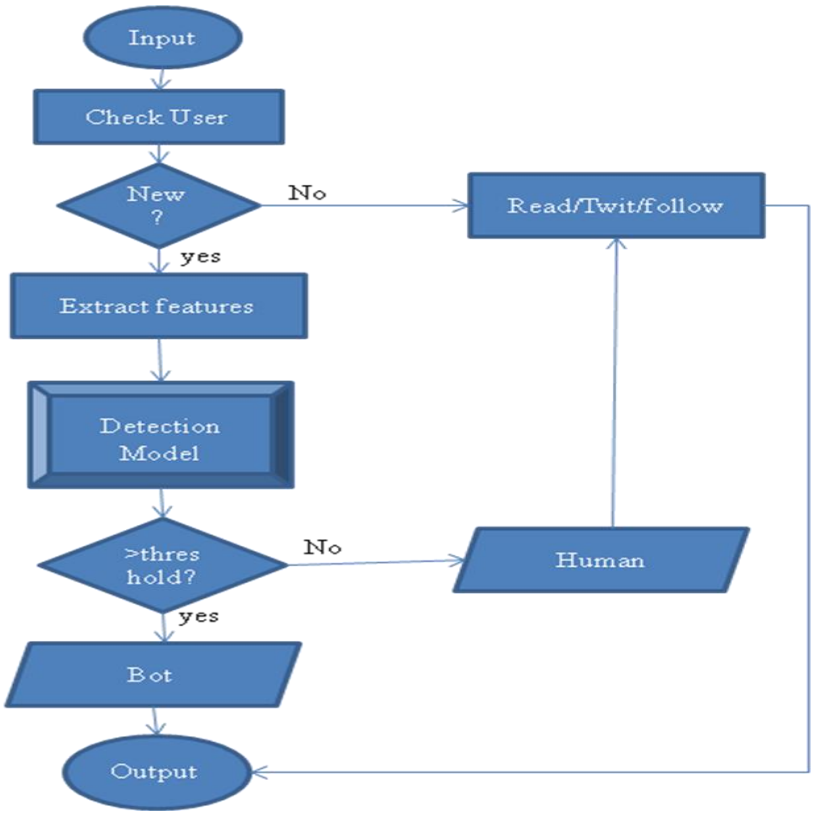


Fig 2: Flow chart

3.3. Conceptual Framework

The intrusion prevention model interfaces between the new user, the social media platform and the human user. Before any social interaction in the virtual space, the human user is expeted to verify the new user to ensure they are not bad bots. To do this, the prevention model can be used to check the status of the new user to identify the class of user it belong. This is achieved using the detection model. Firstly, the web crawler extract the new user account features from the social media platform. This feature dataset will be passed to the detection engine for processing. The output of the processing is user classification as either social bot or human user depending on the threshold assertained by the detection engine. A notification message is communicated to the user to enable the user to determine the level of trust to accord to the new user.

The high level view of the proposed system is presented in figure 4. The social media user triggers a request either POST or GET request to the social media server, the request handler which is the social media platform crawler generate a dataset of account-level features of the user which is passed to the detection layer for analysis and categorization. If the percentage of the likelihood of the account falls below the defined threshold, the user is classified as a social bot, else, the user is classified as a normal human user. The communication of the message is passed to the user through the API for decision making by the social media user that wants to initiate connection with the new user.

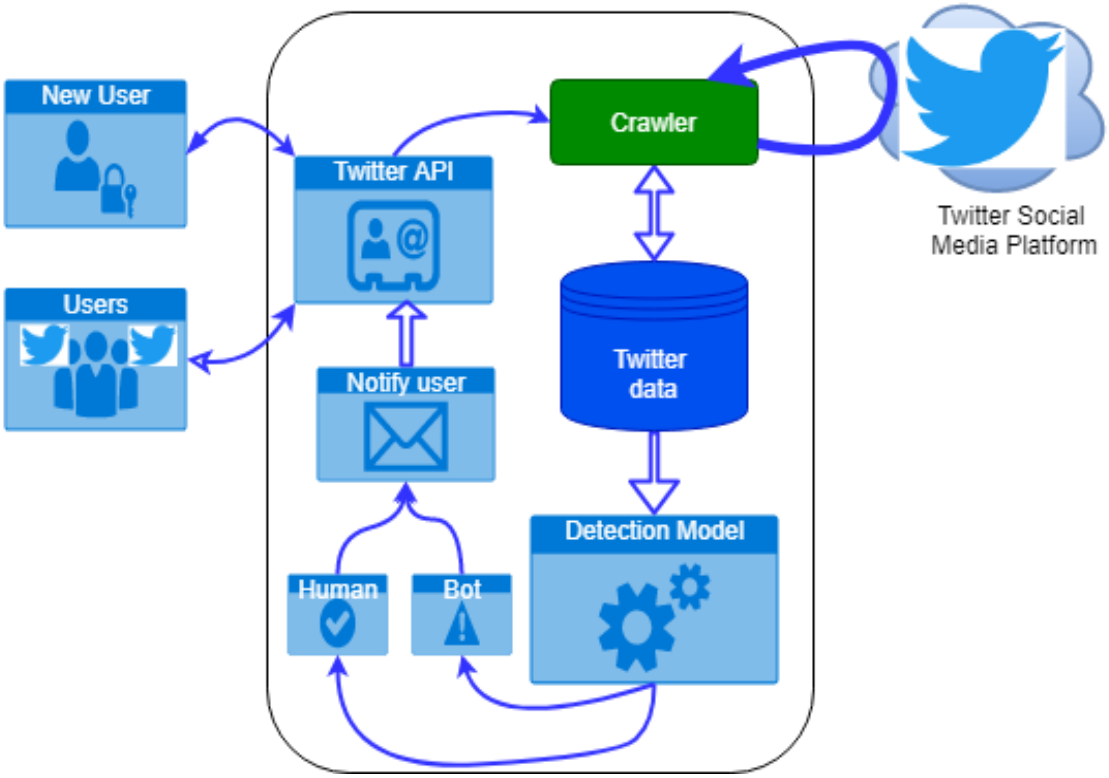


Fig 3: Conceptual model

3.4. Performance Measurement

To evaluate the performance of the model, four standard indicators will be used to evaluate the performance of the model. Machine learning models according to [23] has standard performance indicators for measuring its performance. They are: True Negative (TN), True Positive (TP), False Negative (FN), and False Positive (FP).

i) True Positive (TP): is when the social media account is predicted to belong to a human user or social bot class and it actually does belong to that class.
ii) False Positive (FP) is when the social media account is predicted to belong to a human user or social bot class and it actually does not belong to that class.
iii) True Negative (TN) is when the social media account is predicted not to belong to a human user or social bot class and it actually does not belong to that class.
iv) False Negative (FN) is when the social media account is predicted not to belong to a human user or social bot class and it actually does belong to that class.

Other evaluation parameters are precision, accuracy, recall, and F1 score [23] [24][25]. The f-score is used to weigh the overall performance of a developed machine learning model, accuracy is the number of correctly predicted values out of the total prediction sample space.
Precision is the number of true predictions that were positive against the true positives with the false positives. They are defined by equation (1) - (4).

$$Accuracy = \frac{\text{correct predictions}}{\text{all predictions}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 1$$

$$precision = \frac{\text{true positives}}{\text{true positives} + \text{false positives}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 2$$

$$recall = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3$$

$$F_\beta = (1 + \beta^2 \frac{\text{precision.recall}}{(\beta^2.precision) + recall} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 4$$

4. CONCLUSIONS

The wide scale usage of social media that has attracted the activities of hackers must be securely protected against the malicious activities of hackers that use bad social bot to attack naïve users. Therefore an intelligent intrusion prevention system will greatly enhance the enormous benefits available on the social media network platforms. The research work will design an intelligent intrusion prevention framework that will prevent attacks through bad social bots on social media network platforms. Several attacks which can be launched by hackers using bad social bots will be proactively averted even before they connect virtually with any social media user.

ACKNOWLEDGEMENTS

The authors would like to thank the PG board of the department of Computer Science University of Nigeria for their constructive criticism on this research work.

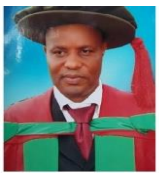
REFERENCES

[1] S. Kudugunta and E. Ferrara, "Deep Neural Networks for Bot Detection," Inf. Sci. (Ny)., vol. 467, pp. 312–322, 2018.
[2] A. El Azab, A. M. Idrees, M. A. Mahmoud, and H. Hefny, "Fake Account Detection in Twitter Based on Minimum Weighted Feature set," Int. J. Comput. Inf. Eng., vol. 10, no. 1, pp. 13–18, 2016.
[3] P. G. Efthimion, S. Payne, and N. Proferes, "Supervised Machine Learning Bot Detection Techniques to Identify Social Twitter Bots," SMU Data Sci. Rev., vol. 1, no. 2, 2018.
[4] A. E. Omolara, A. Jantan, O. I. Abiodun, and E. Etuh, "Balancing Security and Application Functionality in Cloud-based Applications : A Survey," Int. J. Electr. Eng., vol. 12, no. 1, pp. 1–20, 2019.
[5] E. Etuh, A. A. Obiniyi, O. S. Yusuf, and O. A. Akinyele, "Design and Implementation of a Secure Online Electronic Transaction (SOET) System for a Cashless Society," J. Comput. Sci. Appl., vol. 5, no. 2, pp. 83–89, 2017.
[6] Z. Umar and E. Etuh, "A Framework for Digital Forensic in Joint Heterogeneous Cloud Computing Environment," J. Futur. Internet, vol. 3, no. 1, pp. 1–11, 2019.
[7] A. E. Omolara, A. Jantan, O. I. Abiodun, V. Dada, H. Arshad, and E. Emmanuel, "A Deception Model Robust to Eavesdropping over Communication for Social Network Systems," no. Im, pp. 1–21, 2019.
[8] E. Etuh, F. S. Bakpo, and E. A. H, "Social Media Network Attacks and Their Preventive Mechanisms: A Review," Comput. Sci. Inf. Technol. (CS IT), vol. 11, no. 4, pp. 59–72, 2021.
[9] M. S. Rahman, S. Halder, M. A. Uddin, and U. K. Acharjee, "An efficient hybrid system for anomaly detection in social networks," Cybersecurity, vol. 4, no. 10, pp. 1–11, 2021.
[10] A. Singhal and S. Jajodia, "Data warehousing and data mining techniques for intrusion detection systems," Distrib Parallel Databases, vol. 20, pp. 149–166, 2006.
[11] G. N. Prabhu, K. Jain, N. Lawande, Y. Zutshi, R. Singh, and J. Chinchole, "Network Intrusion Detection System," Int. J. Eng. Res. Appl., vol. 4, no. 4, pp. 69–72, 2014.

[12] R. A. Jamadar, "Network Intrusion Detection System Using Machine Learning," Indian J. Sci. Technol., vol. 11, no. 48, pp. 1–6, 2018.
[13] L. Rovito, L. Bonin, L. Manzoni, and A. De Lorenzo, "An Evolutionary Computation Approach for Twitter Bot Detection," Appl. Sci., vol. 5915, no. 12, pp. 1–25, 2022.
[14] V. Calleja-solanas et al., "Twitter bot detection using supervised machine learning," J. Phys. Conf. Ser., vol. 1950, no. 2021, pp. 1–11, 2021.
[15] D. Kosmajac and V. Keselj, "Twitter Bot Detection using Diversity Measures," in 3rd International Conference on Natural Language and Speech Processing, 2019, pp. 1–8.
[16] F. Wei and U. T. Nguyen, "Twitter Bot Detection Using Bidirectional Long Short-term Memory Neural Networks and Word Embeddings," in First IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications, 2019, pp. 101–109.
[17] T. . Subramaniam and B. Deepa, "A Scrutiny to Attack Issues and Security Challenges in Cloud Computing," Int. J. Ambient Syst. Appl., vol. 4, no. 1, pp. 1–10, 2016.
[18] S. Wen and Y. Sun, "An Intelligent and Data-Driven Mobile Volunteer Event Management Using Machine Learning and Data Analytics," Int. J. Ambient Syst. Appl., vol. 9, no. 3, pp. 1–17, 2021.
[19] O. Logvinov, "Standard for an Architectural Framework for the Internet of Things (IoT)," 2021. .
[20] H3C, "Social Media Attacks," USA, 2020.
[21] K. Musial and P. Kazienko, "Social networks on the Internet," World Wide Web, pp. 31–72, 2012.
[22] H. Vora, J. Kataria, D. Shah, and V. Pinjarkar, "Intrusion Detection System for College ERP System," J. Res., vol. 03, no. 02, pp. 69–72, 2017.
[23] B. Shanmugam and N. B. Idris, "Artificial Intelligence Techniques Applied To Intrusion Detection," in Proceedings of the Postgraduate Annual Research Seminar, 2005, pp. 285–287.
[24] E. Alpaydin, Introduction to Machine Learning, Second. London, England: The MIT Press, 2010.
[25] J. Davis and M. Goadrich, "The Relationship Between Precision-Recall and ROC Curves," in 23rd International Conference on Machine Learning, 2006, pp. 233–240.
[26] R. J. Santos, J. Bernardino, and M. Vieira, "DBMS Application Layer Intrusion Detection for Data Warehouses," in Building sus- tainable information systems., 2013.
[27] A. Arora and A. Gosain, "Intrusion Detection System for Data Warehouse with Second Level Authentication," Int. J. Inf. Technol., vol. 13, pp. 877–887, 2021.

AUTHORS

Emmanuel Etuh is a lecturer in the department of Mathematics, Statistics, and Computer Science at Kwararafa University, Wukari, Nigeria and currently pursuing a PhD degree in Computer Science at the University of Nigeria, Nsukka. He obtained his first degree certificate in Computer Science from Kogi State University, Anyigba in 2009 and an MSc degree in Computer Science from Ahmadu Bello University, Zaria in 2014, His research interests include Artificial Intelligence, Cyber Security, and Software Engineering.



Okereke George Emeka is a senior Lecturer/Researcher, Computer Science Department, University of Nigeria, Director, Computing Centre, Former Head of Department, Computer Science, University of Nigeria. He obtained a Bachelor of Engineering (Hons.) in Computer Science & Engineering from Enugu State University of Science and Technology and a Master of Science degree in Computer science from University of Nigeria. His PhD is in Digital Electronics & Computing from Electronic Engineering Department of University of Nigeria. He joined the services of University of Nigeria in 1998 as a lecturer in Computer Science Department and is currently a Senior Lecturer. Head of Department from 2017 to 2019. His research interest is in Network security, web security, computer forensics, electronic transfers and security, web design and computer architecture/design. George is married with six children.

Dr. Ir. Engr. (Mrs.) Deborah Uzoamaka Ebem is a lecturer in the department of Computer Science, University of Nigeria, Nsukka. Deborah is a native of Ugbo In Awgu Local Government Area of Enugu State. She received a B.Engr. degree from Anambra State University of Technology (ASUTECH) and a postgraduate diploma in management from University of Nigeria, Nsukka. She also received master's degrees in Computer Science and Engineering and Computer Engineering from Enugu State University of Science (ESUT) and the Technical University Delft, The Netherlands, respectively. She further holds a PhD in Computer Science from Ebonyi State University, Abakaliki, Nigeria. She was a Research Fellow and Scholar at Massachusetts Institute of Technology (MIT), Cambridge USA.

Bakpo Francis S is a Professor in the Department of Computer Science, University of Nigeria, Nsukka. He joined the Department of Computer Science, University of Nigeria, Nsukka as a Corp member in 1995, retained by the Department in 1996 as lecturer II and progressed to Professor in 2010. He received his Master's degree in Computer Science and Engineering from Kazakh National Technical University, Almaty (formerly, USSR) in 1994 and Doctorate degree in Computer Engineering in 2008 from Enugu State University of Science and Technology, Agbani. Area of Specialization include: computer architecture, computer communications network, Artificial neural network, intelligent software agents and Petri nets theory and applications.