

An AI Approach to Mitigating Online Fraud, Phishing as a Case Study

¹Prof. B. O. Omijeh, ²Eberechi Victor, ³Aji Jacob Onu

¹NCC Professorial Chair, University of Port Harcourt, Rivers State, Nigeria

²University of Port Harcourt, Port Harcourt, Nigeria

³Centre for Information and Telecommunication Engineering, University of Port Harcourt, Rivers State, Nigeria

Abstract - This form of fraud called phishing is a form of threat that involves exploiting users and organizations through malicious or deceptive emails and web URLs. In this study, explored are some of the application of artificial intelligence data analysis and machine learning techniques applied in the detecting, preventing and mitigating the prevalence attacks particularly phishing here, diverse dataset have been leveraged to develop high accuracy models that can help detect legitimate and malicious emails or legitimate and malicious URL, here we extract features like certain keywords, certain behaviors, URL blacklisting, URL shortening services, etc NLP (natural language processing) methods like TF-IDF are used to enhance model accuracy and precision algorithms like random forest decision tree extra tree and XGboost are used. Our result demonstrates high accuracy and how important the use of multiple algorithms can be. This research considers the potential of AI-driven solutions in mitigating fraud with particular respect to phishing.

Keywords: Phishing URL, Phishing Email, Algorithm, TF-IDF, Classifier.

I. INTRODUCTION

Phishing attacks have emerged as a significant cybersecurity threat, targeting individuals and organizations to extract sensitive information, financial data, or credentials. This kind of attack is such that malicious emails and URL appear deceptively as legitimate thereby used to steal sensitive information from users or organization. This use of advanced social engineering concepts has made traditional system.

1.1 Statement of the Problem

Despite the wide relinquishment of cybersecurity measures, phishing remains one of the leading causes of data breaches and fiscal losses. Being discovery systems frequently calculate on stationary rule- grounded approaches or blacklists, which are ineffective against fleetly changing phishing tactics and recently generated vicious URLs. These traditional styles struggle to manage with obfuscation ways,

similar as sphere masquerading and URL shortening, or socially finagled content designed to bypass discovery. This exploration addresses this gap by exploring the operation of ML and AI ways to make robust models for phishing dispatch and URL discovery, enabling better protection against evolving cyber pitfalls. <https://nairametrics.com/2024/04/12/nigeria-ranks-5th-in-global-cybercrime-index/>. According to Nairametrics, Nigeria has been ranked 5th in a global report on sources of cybercrime conditioning, coming behind Russia, which ranked number one, and Ukraine, China, and the United States, which enthralled the alternate, third, and fourth positions independently. https://guardian.ng/business-services/cybercrime-rises-as-phishing-successes-174-in-nigeria-438-in-kenya/#google_vignette According to the guarduaian news, cybercrime in Nigeria and Kenya with respect to phishing swindles hit 174 percent and 438 percent independently, (https://guardian.ng/business-services/cybercrime-rises-as-phishing-successes-174-in-nigeria-438-in-kenya/#google_vignette). Kaspersky, which revealed this, in its new analysis history, said attacks related to data loss pitfalls(phishing and swindles social engineering) increased significantly in Africa in Q2 2022 in comparison with the former quarter. Kaspersky says its security results detected phishing attacks in Africa in Q2. <https://www.premiumtimesng.com/news/more-news/649110-phishing-other-cyber-attacks-increase-in-nigeria-others.html?tztc=1>.

The amount lost to fraud in Nigeria (N'billion)

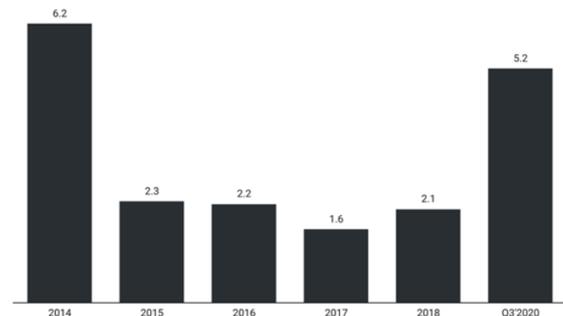


Figure 1: showing the amount lost to fraud in recent years

Phishing attacks that were detected in Kenya in the third quarter of 2023 increased by 32 per cent compared to the second quarter of 2023 and by 12 per cent compared to third quarter of 2022. In Nigeria, there was a 12 per cent increase in phishing attack findings in Q3 2023 compared to Q2; still, compared to Q3 2022, the number of phishing detects dropped by 8 percent.

According to the Nigeria Inter-Bank Settlement System Plc(NIBSS), in the first nine months of 2020, fraudsters tried 46,126 attacks, and they were successful on 41,979 occasions, 91 percent of the time.([https:// guardian.ng/ business-services/ cybercrime- rises as- phishing- successes-174-in-nigeria-438-in-kenya/#google_vignette](https://guardian.ng/business-services/cybercrime-rises-as-phishing-successes-174-in-nigeria-438-in-kenya/#google_vignette))

1.2 Objectives

With the rise in online fraud in Nigeria and beyond, this paper aims to actualize the following objectives:

- Gather datasets of malicious emails and URLs.
- Analyze this data and chose an appropriate algorithm.
- Train the the data and develop a model that detects phishing emails and URLs.

1.3 Significance of the Projects

Phishing is an attack that continually poses threat to our daily life in the digital spaces. This project is crucial due to its proposed potential to detect and mitigate this kind of threat. By leveraging machine learning (ML) and artificial intelligence (AI), the project aims to develop at detection system capable of identifying phishing emails and URLs with high accuracy. This will enhance protection against financial fraud, identity theft, and data breaches, mitigating the substantial economic and reputational damages caused by phishing attacks. Phishing is a global issue, impacting organizations of all sizes and industries. The outcomes of this project have the potential for widespread adoption, fostering safer online ecosystems and reducing the overall prevalence of phishing-related crimes.

II. RELATED WORKS

The field of phishing detection has evolved significantly, leveraging advancements in machine learning (ML) and artificial intelligence (AI) to develop robust detection systems. This review synthesizes key contributions to phishing email and URL detection, focusing on the methodologies, datasets, and algorithms employed.

Ahmed and Abdullah (2016) considers the effectiveness of real-time phishing website detection by analyzing the features of web content [1]. Also, Akinyelu and Adewumi

(2014) used the Random Forest algorithm to classify phishing emails, thereby looking at its potential to handle large dataset with high accuracy [2]. Alauthman et al. (2019) explored ML techniques for phishing detection and mitigation, emphasizing the importance of hybrid methods [3] . In a similar study, Ansari et al. (2022) proposed AI-driven algorithms for phishing prevention, particularly focusing on adaptive models that can identify evolving attack patterns.[4]

Elsadig et al. (2022) introduced an intelligent phishing URL detection system based on BERT for feature extraction, achieving superior performance in terms of accuracy and scalability [6]. Fang et al. (2019) improves this by applying an improved RCNN model with attention mechanisms for phishing email detection, thereby allowing for the system to capture concepts from texts accurately [7] In addition to detection systems, several researchers have examined the role of interpretability and user feedback in phishing detection. Psychoula et al. (2021) emphasized the need for explainable ML in fraud detection[13]. Gupta et al. (2020) combined user feedback with a Naïve Bayesian approach, illustrating how user-centric designs can enhance the accuracy of email classification systems [8].

The use of semi-supervised learning for phishing webpage detection, as demonstrated by Li et al. (2013), highlighted the potential of leveraging partially labeled data to improve model performance [12]. This is particularly relevant in scenarios where acquiring fully annotated datasets is challenging. Kumar et al. (2023) further explored fake URL detection using LSTMs, emphasizing the importance of sequence-based models in capturing temporal and structural dependencies in phishing URLs.[11]

Jain et al. (2021) reviewed security and privacy issues in online social networks, underscoring the role of social engineering in phishing and the need for proactive detection mechanisms [10].The combined insights from these studies underline the growing sophistication of phishing attacks and the corresponding need for adaptive, intelligent, and interpretable detection systems. This literature review forms the basis for developing hybrid models that integrate advanced ML algorithms with user feedback mechanisms to address the limitations of current solutions. The application of machine learning (ML) and natural language processing (NLP) in phishing detection has been extensively researched, with various approaches focusing on analyzing URLs and email content. Sahingoz et al. (2018) presented a comprehensive ML-based framework for phishing detection, leveraging URL features such as lexical patterns and token structures [15]. Their work demonstrated the effectiveness of ensemble learning techniques in distinguishing between phishing and

legitimate URLs with high accuracy. Similarly, Tang and Mahmoud (2021) conducted a detailed survey of ML-based solutions for phishing website detection, emphasizing the evolution of feature engineering and model architectures over time [18].

Feature selection remains a critical challenge in phishing detection. Zareapoor and R (2015) examined feature extraction and selection techniques in the context of text classification for phishing emails, demonstrating how optimized feature sets can enhance model performance[20].The use of hybrid models has also been explored. Verma et al. (2020) discussed how combining traditional ML algorithms with advanced NLP techniques can yield better results for phishing email detection [19].

These studies and research showcase the diversity of approaches and challenges, and the need for advancement.

III. RESEARCH AND METHODOLOGY

3.1 Overview

This section focuses on the proposed algorithms and methods in used in mitigating online frauds which includes Phishing emails and URL. Data collected were majorly secondary data from sources like, kaggle, phishtanketc.

Here, Natural language processing is introduced for text classification, models are trained to either classify as phishing mail or legitimate email based on their features like phrases, keywords and sentence structure, emotional tone is also analyzed in order to identify urgent or threatening message. Algorithms like Naive Bayes, Random forest and Decision tree were used. For the Phishing URL, the concept of URL blacklisting was used where we maintained a database of malicious URLs and check incoming URLs against this list, Domain age and reputation was also analyzed to identify malicious websites.

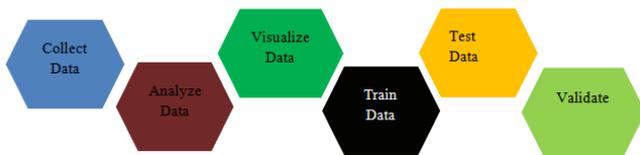


Figure 3.1: Showing the overall workflow of the research

General tools used In development of the model include the following are Python, Scikit Learn, Google Colab.

3.2 Phishing Email Scam

This section aims to propose a detection model for mitigation of phishing email scam Figure 3.0 shows the

overall process of our proposed method. this proposed model has was trained on the Naive Bayes Algorithm for better accuracy the algorithm used in this case is the Naive Bayes specifically the multinomial naive Bayes. Here textual features like keyword (e.g payment, urgent, confirm account) are extracted from the datasets, emotional tone of the email is also analyzed, syntactic features like grammar and punctuation, and language structure are also extracted like. The algorithm creates relevant features from the extracted features and the model is trained on the prepared datasets using the extracted features as input.

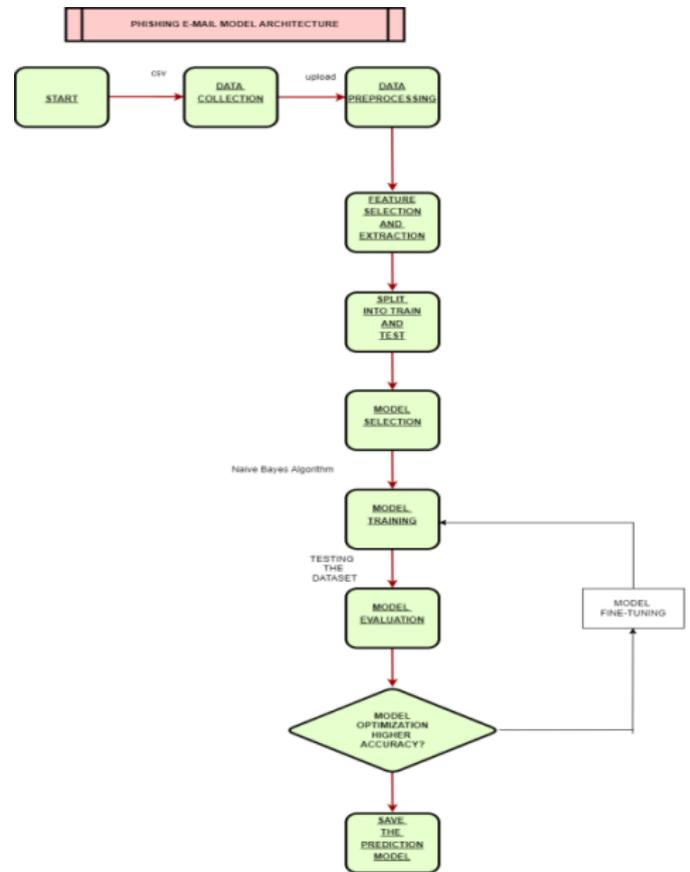


Figure 3.2: Flowchart for the Phishing email detection model

3.2.1 Dataset

In the dataset we have used secondary data of emails contain two categories of safe mails and phishing emails. We collected 18649 safe and phishing mail. The dataset consists of 50518 Phishing Email and 50467 Safe emails samples.

3.2.2 Data Pre-processing

After collecting data, our next step was to pre-process it further. Dataset without pre-processing is shown below. in pre-processing the dataset, the following tools were used are, Numpy, Pandas, Seaborn, Nltk, Re.

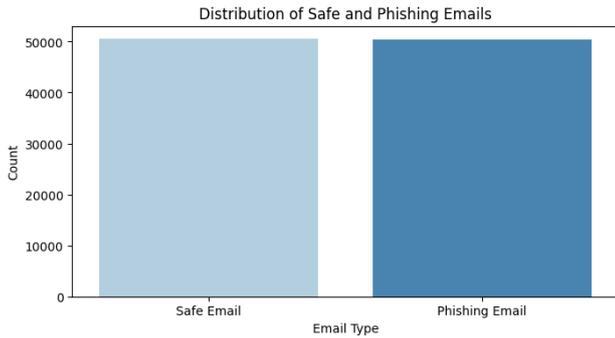
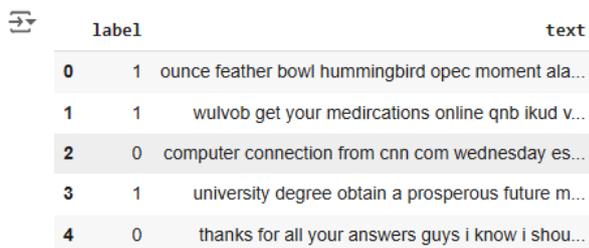


Figure 3.3: Showing datasets distribution

This is a very crucial step in data analysis an machine learning as it helps n increasing accuracy. Here, the models label was encoded as ‘1’ and ‘0’ which indicates email is classified as phishing and email is legitimate respectively. Some of the activities in the data pre-processing stage are removing hyperlinks, removing lower cases, tokenization, removing special characters, removing stop words and punctuation and finally stemming. This was made possible by the use of the function porter stemmer which is a function from the nltk module.

After removing special characters and tokenization, another technique used was the TF-IDF technique. TF-IDF (Term Frequency-Inverse Document Frequency). TF-IDF is a very technique in NLP that retrieves information from a document and state how important that particular word is for that document. This technique initially transforms all the words into individual vectors. Then it retrieves features that work with vectors. The TF part calculates the frequency of a particular term against the total word count in total document and the IDF (Inverse Document Frequency) evaluates the weight of that term in the document. After that, with both TF and IDF values we count our final TF-IDF values for all the vectors. Then, we created a new column with extracted features from every column that has a text value: TF-IDF (Term Frequency-Inverse Document Frequency) assigns weights to terms based on their frequency within an email compared to the entire dataset, helping highlight important keywords that might signal phishing.



| | label | text |
|---|-------|---|
| 0 | 1 | ounce feather bowl hummingbird opec moment ala... |
| 1 | 1 | wulvob get your medircations online qnb ikud v... |
| 2 | 0 | computer connection from cnn com wednesday es... |
| 3 | 1 | university degree obtain a prosperous future m... |
| 4 | 0 | thanks for all your answers guys i know i shou... |

Figure 3.4: Showing the dataset snippet after preprocessing

3.2.3 Algorithm and Model development

In this research we have used the Naive Bayes algorithms to predict whether an email is safe or unsafe. We experimented with some latest algorithms that have not been used frequently before to determine how they work on our dataset.

The model was trained by the train test split function from the SK-learn framework with a random state of 42 and a train test ratio of 85:15 (85% to inform 15%)

Naive Bayes Algorithm: A simple model often effective for text classification due to its assumption of independence among features.

3.3 Phishing URL Scam

This section of the models aims to develop a model that targets phishing URL. We gathered data containing URL links divided in the following order benign, defacement, phishing and malware respectively. The figure below shows a snippet of the dateset before pre-processing.



| | url | type |
|---|---|------------|
| | br-icloud.com.br | phishing |
| 1 | mp3raid.com/music/krizz_kaliko.html | benign |
| 2 | bopsecrets.org/rexroth/cr/1.htm | benign |
| 3 | http://www.garage-pirenne.be/index.php?option=... | defacement |
| 4 | http://adventure-nicaragua.net/index.php?optio... | defacement |

Figure 3.5: Showing the dataset before preprocessing

3.3.1 Data Pre-processing

The first step after gathering the data was to clean it by checking for null values, next was to visualize and make caparisons to our target column, the chart below show the comparison of our target column.

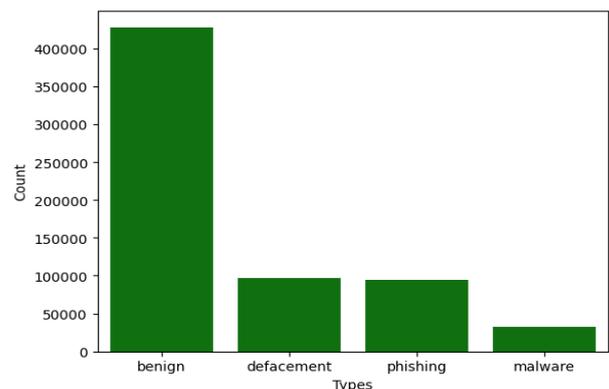


Figure 3.6: Showing the different classes of phishing urls

The next step was to encode the target categories benign, defacement, phishing, and malware with values 0, 1, 2, 3 respectively. The next step was to handle special characters and check for the use of URL shortening services, and finally we checked if the URL had secure socket layer certificate. After the above steps in dataset was then fed into the algorithm for training.

3.3.2 Algorithm and Model development

Several Machine learning algorithms were used to train the dataset and predict the result, we used some of the latest algorithms, this is crucial for better accuracy. These algorithms include Decision tree classifier, Adaboost classifier, Random forest classifier, SGD classifier, Extra tree classifier.

3.3.2.1 Decision Tree Classifier

A decision tree is a flowchart-like structure used to make decisions or predictions. It consists of nodes representing decisions or tests on attributes, branches representing the outcome of these decisions, and leaf nodes representing final outcomes or predictions.

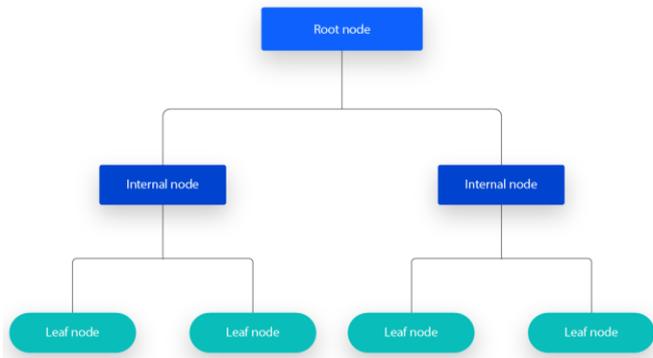


Figure 3.7: Showing the decision tree classifier

3.3.2.2 Adaboost Classifier

Adaboost classifier can be used in conjunction with many types of learning algorithm to improve performance. Usually, AdaBoost is presented for binary classification, although it can be generalized to multiple classes or bounded intervals of real values. AdaBoost is adaptive in the sense that subsequent weak learners (models) are adjusted in favor of instances misclassified by previous models. In some problems, it can be less susceptible to overfitting than other learning algorithms.

$$F_T(x) = \sum_{t=1}^T f_t(x)$$

Figure 3.8: Adaboost classifier

3.3.2.3 Stochastic Gradient Descent (SGD) is a simple yet very efficient approach to fitting linear classifiers and regressors under convex loss functions such as (linear) Support Vector Machines and Logistic Regression.

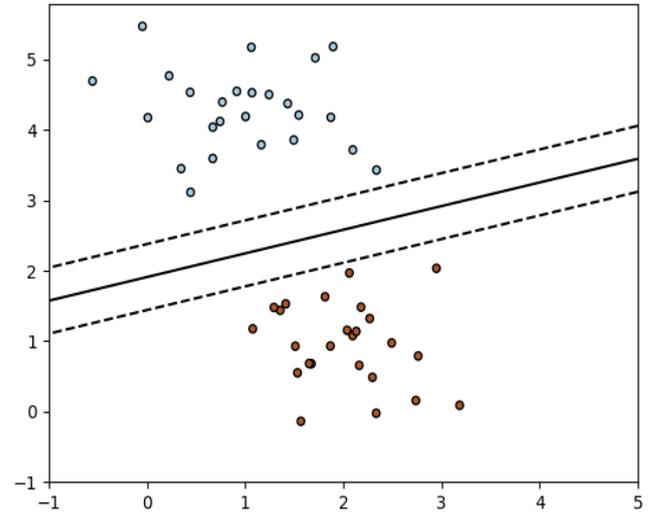


Figure 3.9: Showing the SGD classifier

3.3.2.4 Extremely Randomized Trees Classifier (Extra Trees Classifier) In concept, it is very similar to a Random Forest Classifier and only differs from it in the manner of construction of the decision trees in the forest. Each Decision Tree in the Extra Trees Forest is constructed from the original training sample.

IV. RESULTS AND DISCUSSIONS

4.1 Data Visualization

Visualizing the data is an easy way to understand the data and patterns through the use of charts, plots and heat maps. It refers to the graphical representation of the data. For this paper we have used many tools and libraries in python namely matplotlib, seaborn etc.

4.2 Heatmap

Below is the heat map using the seaborn library. It is a colored representation of the correlation between the columns in the Phishing email datasets. Here the blocks more correlated get cold colors while those less correlated get lighter colors.

model performance. The integration of features like emotional tone analysis and domain reputation verification, which significantly bolstered the models' detection capabilities. The models perform with a peak accuracy of about 95% for phishing email detection and 91% for phishing URL detection.

REFERENCES

- [1] Ahmed, A. A., & Abdullah, N. A. (2016). Real time detection of phishing websites. 2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 1–6. <https://doi.org/10.1109/iemcon.2016.7746247>
- [2] Akinyelu, A. A., & Adewumi, A. O. (2014). Classification of Phishing Email Using Random Forest Machine Learning Technique. *Journal of Applied Mathematics*, 2014, 1–6. <https://doi.org/10.1155/2014/425731>
- [3] Alauthman, M., Almomani, A., Alweshah, M., Omoush, W., & Alieyan, K. (2019). Machine Learning for Phishing Detection and Mitigation. In *CRC Press eBooks* (pp. 48–74). <https://doi.org/10.1201/9780429504044-2>
- [4] Ansari, M. F., Panigrahi, A., Jakka, G., Pati, A., & Bhattacharya, K. (2022). Prevention of Phishing attacks using AI Algorithm. *Phishing*, 1–5. <https://doi.org/10.1109/odicon54453.2022.10010185>
- [5] Chawla, A., & Kohli, S. S. (2022). Phishing Site Detection Using Artificial Intelligence. In *Lecture notes in electrical engineering* (pp. 667–681). https://doi.org/10.1007/978-981-19-5037-7_48
- [6] Elsadig, M., Ibrahim, A. O., Basheer, S., Alohal, M. A., Alshunaifi, S., Alqahtani, H., Alharbi, N., & Nagmeldin, W. (2022). Intelligent Deep Machine Learning Cyber Phishing URL Detection Based on BERT Features Extraction. *Electronics*, 11(22), 3647. <https://doi.org/10.3390/electronics11223647>
- [7] Fang, Y., Zhang, C., Huang, C., Liu, L., & Yang, Y. (2019). Phishing Email Detection Using Improved RCNN Model with Multilevel Vectors and Attention Mechanism. *IEEE Access*, 7, 56329–56340. <https://doi.org/10.1109/access.2019.2913705>
- [8] Gupta, A., Palwe, S., & Keskar, D. (2020). Fake Email and Spam Detection: User Feedback with Naives Bayesian Approach. In *Algorithms for intelligent systems* (pp. 41–47). https://doi.org/10.1007/978-981-15-0790-8_5
- [9] Gupta, B. B., Arachchilage, N. a. G., & Psannis, K. E. (2017). Defending against phishing attacks: taxonomy of methods, current issues and future directions. *Telecommunication Systems*, 67(2), 247–267. <https://doi.org/10.1007/s11235-017-0334-z>
- [10] Jain, A. K., Sahoo, S. R., & Kaubiyal, J. (2021). Online social networks security and privacy: comprehensive review and analysis. *Complex & Intelligent Systems*, 7(5), 2157–2177. <https://doi.org/10.1007/s40747-021-00409-7>
- [11] Kumar, S. a. B. B. V. K. P. N. P. R. (2023). Fake URL Detection Using LSTM. Zenodo (CERN European Organization for Nuclear Research). <https://doi.org/10.5281/zenodo.7794116>
- [12] Li, Y., Xiao, R., Feng, J., & Zhao, L. (2013). A semi-supervised learning approach for detection of phishing webpages. *Optik*, 124(23), 6027–6033. <https://doi.org/10.1016/j.ijleo.2013.04.078>
- [13] Psychoula, I., Gutmann, A., Mainali, P., Lee, S. H., Dunphy, P., & Petitcolas, F. (2021). Explainable Machine Learning for Fraud Detection. *Computer*, 54(10), 49–59. <https://doi.org/10.1109/mc.2021.3081249>
- [14] Raghavan, P., & Gayar, N. E. (2019). Fraud detection using machine learning and deep learning. *Fraud Detection*, 334–339. <https://doi.org/10.1109/iccike47802.2019.9004231>
- [15] Sahingoz, O. K., Buber, E., Demir, O., & Diri, B. (2018). Machine learning based phishing detection from URLs. *Expert Systems With Applications*, 117, 345–357. <https://doi.org/10.1016/j.eswa.2018.09.029>
- [16] Salloum, S., Gaber, T., Vadera, S., & Shaalan, K. (2021). Phishing Email Detection Using Natural Language Processing Techniques: A Literature Survey. *Procedia Computer Science*, 189, 19–28. <https://doi.org/10.1016/j.procs.2021.05.077>
- [17] Sonowal, G. (2020). Phishing Email Detection Based on Binary Search Feature Selection. *SN Computer Science*, 1(4). <https://doi.org/10.1007/s42979-020-00194-z>
- [18] Tang, L., & Mahmoud, Q. H. (2021). A Survey of Machine Learning-Based Solutions for Phishing Website Detection. *Machine Learning and Knowledge Extraction*, 3(3), 672–694. <https://doi.org/10.3390/make3030034>
- [19] Verma, P., Goyal, A., & Gigras, Y. (2020). Email phishing: text classification using natural language processing. *Computer Science and Information Technologies*, 1(1), 1–12. <https://doi.org/10.11591/csit.v1i1.p1-12>
- [20] Zareapoor, M., & R, S. K. (2015). Feature Extraction or Feature Selection for Text Classification: A Case Study on Phishing Email Detection. *International Journal of Information Engineering and Electronic Business*, 7(2), 60–65. <https://doi.org/10.5815/ijieeb.2015.02.08>

[21] <https://nairametrics.com/2024/04/12/nigeria-ranks-5th-in-global-cybercrime-index/>

[22] https://guardian.ng/business-services/cybercrime-rises-as-phishing-successes-174-in-nigeria-438-in-kenya/#google_vignette.

AUTHORS BIOGRAPHY



Prof. B. O. Omijeh is a distinguished digital educator, registered engineer, seasoned ICT professional and erudite scholar with a strong passion for academia, innovation, and faith. He holds the Professorial Chair on ICT & Telecommunications at the University of Port Harcourt (UNIPORT), sponsored by the Nigerian Communications Commission (NCC).



Victor Eberechi is a passionate and driven graduate with a Bachelor of Engineering (B.Eng.) in Electrical/Electronic Engineering from the University of Port Harcourt. With a strong foundation in both theoretical and practical aspects of engineering, Victor has developed hands-on expertise in embedded systems, IoT development, web development, blockchain technology (Solidity, Web3.js), and data science (Python, SQL, Jupyter Notebooks). Victor is also skilled in tools and platforms like MATLAB/Simulink, Scilab/Xcos.



Aji Jacob Onu holds MSc and PhD. In Information and Telecommunications Engineering from Centre for information and telecommunications Engineering, University of Port Harcourt, Rivers State, Nigeria. His research areas includes Telecommunications Network, Frequency reuse, Machine learning (ML), 5G optimization etc. He is a registered member of National Association of Technological Engineers (NATE), a certify Lead Auditor FMS (ISO 41001), Cisco Certified Network Associate (CCNA) and Facilities Management Professional in Telecommunications Domain.

Citation of this Article:

Prof. B. O. Omijeh, Eberechi Victor, & Aji Jacob Onu. (2025). An AI Approach to Mitigating Online Fraud, Phishing as a Case Study. *International Current Journal of Engineering and Science - ICJES*, 4(7), 18-25. Article DOI: <https://doi.org/10.47001/ICJES/2025.407002>
