Machine Learning Model to Mitigate Fake Bank Alert Phishing Fraud in Nigeria: The Initial Investigation

Ebiesuwa Seun and Tepede Dipo

Abstract:

This paper suggests a conceptual model based on machine learning to combat the persistent threat of Fake Bank Alert (FBA) phishing attacks in Nigeria's Internet Banking (IB) system. FBA fraud targets small businesses by masquerading as a bank transaction alert, leading to financial losses and brand erosion. To address this, an extensive literature review was conducted, identifying gaps and informing the development of a model with concrete future recommendations.

Keywords – Internet Banking, Phishing Attacks, Fake Bank Alert (FBA), Machine Learning, Security, Nigeria, Detection, Prevention, URL Characteristics, Decision Tree Ensembles

1.0 Introduction

The pervasive use of Internet technology has reshaped daily life and revolutionized global business practices [Khedmatgozar & Shahnazi, (2017)]. Internet Banking (IB), often considered a groundbreaking technological innovation, possesses the potential to reshape the banking landscape by enhancing communication capabilities and transaction efficiency for users [Bharat and Broder, (1998)]. IB's acceptance has grown among banking customers due to its convenience. Still, it has also introduced a set of risks, primarily stemming from its dependence on internet connectivity, which has heightened the susceptibility to online fraud and phishing attacks [Al-Qahtani & Cresci, (2022)], [Gomes *et al*., (2022)].

Phishing Attacks (PAs) represent one of the most formidable information security threats, typically involving tactics that deceive users into disclosing their personal information, altering, or deleting sensitive data, and maliciously disrupting users' resources [Jamil *et al*., (2021)], [Kanhere *et al*., (2018)]. Regrettably, phishing remains a substantial danger to both banking institutions and their clientele [Mridha *et al*., (2017)], with anti-phishing techniques and endeavors thus far falling short in completely thwarting these attacks [Basit *et al*., (2020)], [(He *et al*., 2015)]. Compounding the issue is the existing lack of robust security measures within most IB systems, rendering them vulnerable to PAs [Alsayed and Bilgrami, (2017)]. As such, banking institutions must prioritize safeguarding their users' information from illicit entities seeking to exploit IB accounts for fraudulent activities.

In [Kagantech, (2023)], a Nigerian tech influencer raises awareness about ProBank Fake Bank Alert (FBA) and the dangers of the deceptive practices and fraudulent activities associated with these fake transfer scams. These latest scams often involve tricking individuals into believing they are receiving funds or transfers through SMS or Email messages disguised as a transaction alert from a bank, but it’s a fraudulent scheme designed to steal merchandise from unsuspecting small business owners [Michael, (2022)].

Other Fake Bank Alert applications used to perpetrate different levels of fraud were detailed and comprehensively explained on how they work in [Mitrobe Network, (2022)] [Mitrobe (2022)]. They include the Flash Fund app, Money Prank Pro, and Millionaire Fake Bank Account [Michael, (2022)]. For the successful implementation of this fraud, the perpetrators require the victim’s bank name, bank credit or debit alert format, phone number, and account number which can only be accessed through phishing attacks. Considering these challenges, this paper proposes an ML-based solution to address the persistent threat of phishing attacks in the context of IB [Columbus, (2020)].

2.0 Related Works

A report by Forbes [Columbus, (2020)], clearly advocates utilizing machine learning solutions proves to be an effective strategy for narrowing the disparity between secured nodes and the overall endpoint population. so that Banks, their employees, and their customers can improve their security posture and protect themselves from phishing attacks. Some studies argue that only a technology-based solution is inadequate to ensure the resolution of significant IT security issues [Arachchilage *et al*., (2016)], [Arachchilage and Love (2014)]. They advocate end-user responsibility and well-designed security awareness campaigns targeted at influencing security behavior

[Jansen, (2015)], [Fujita *et al.*, (2015)]. Nonetheless, there is no clarification if assessing phishing avoidance behavior through system trust significantly impacts real-world outcomes [Aribake and Zahurin, (2020)].

There are also other anti-phishing solutions using diverse methods [Yang *et al*., (2019)], [Zhu *et al*., (2020)], such as manual reporting and site analysis heuristics, to develop blacklists Universal Resource Locator (URL) knowledgebase [Balogun *et al*., (2021)]. However, these blacklist anti-phishing solutions fail when new requested URLs are compared with the stored phishing website URLs [Gupta *et al*., (2018)]. Also, some published text condemns blacklisting for its recentness and incorrect evaluation of many malicious websites [Urias, *et al*. (2017)], [Ahammad *et al*., (2022)].

In comparison, ML can discern phishing websites through diverse attributes including domain name, URL composition, website content, and behavioral patterns exhibited by the site [Yang et al., (2019)], [Lokesh and BoreGowda, (2020)]. This makes it more likely to detect newer phishing websites that are not yet blacklisted. Also, Nigerian research reports that an ML-based technique effectively detects legitimate and phishing websites with 98.51% accuracy and a 0.015 false positive rate [Balogun *et al*., (2021)] amongst other positive reports on effective phishing attacks preventive ML studies [Basit *et al*., (2020)], [Volkamer *et al*., (2017)], [Othman and Hassan, (2022)]. Furthermore, a study on using the ML model with blockchain application websites, such as Bitcoin and Ethereum, recorded 98.28% accuracy in detecting blockchain phishing attempts [Joshi et al., (2023)].

Consequently, a structured standardized ML framework was proposed [Bourke, (2019)]. According to the model as depicted in Figure 1, the six steps are not sequentially rigid, but they act as a rough guide to developing the conceptual model.

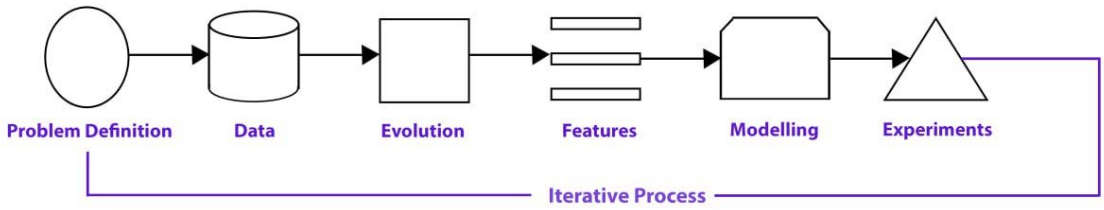

Figure 1 - steps within the data modeling phase [Bourke, (2019)]

(1) Problem Definition:

Detection or classification is probably the most generic and extensively used ML-based solution to preventing phishing [Calzarossa et al., (2023)]. However, some studies prefer to experiment with detection before classification [Mahmoud and Hesham, (2022)] [Basit et al., (2020)], since classification will require better feature determination to enhance accuracy, and expanded datasets facilitate enhanced utilization of Deep Learning (DL) for more precise classification of phishing raids. [Vrbanˇciˇc et al, (2018)] [Shie, (2020)]. Nevertheless, the result of some studies shows that some Deep Learning algorithms, such as Artificial Neuro Networks (ANN), do not produce better performance metrics at classification when compared with some Decision Trees ensembles, such as Random Forest (RF) and Extreme Gradient Boost (XG Boost) [Joshi et al., (2023)], [Amini et al., (2022)].

Similarly, it is important to note that detection and classification are not mutually exclusive, hence, they can both be designed into a Phishing prevention system. Notwithstanding, it is logical that detection is sequenced before classification so that it can be blacklisted, blocked, or reported to the authorities. Then, classification can be used to uncover insight on the types of phishing websites that are being used, and iteratively, to develop more effective detection methods.

(2) Data:

The dataset used for training and testing in ML can be either static (Historic Data), streaming (constantly updating old with new data), or both [Bourke, (2019)]. Ideally, each static structured dataset is equally divided into Benign and Malicious instances of websites when collecting data for either ML phishing detection or classification [Calzarossa *et al*., (2023)], [Chiew et al., (2019):], [Rahman et al., (2020)]. However, in [Bahaghighat et al, (2023)], there was a case of imbalance, and a technique known as SMOTEENN, a combination of SMOTE

(Synthetic Minority Over-Sampling Technique) and Edited Nearest Neighbor (ENN), was used for oversampling imbalanced datasets to balance datasets to improve ML model accuracy. Conversely, some studies, without the aid of a balancing technique, have used imbalance datasets to produce near-perfect performance [Joshi et al., (2023)], [Orunsolu et al, (2022)], [Hota et al., (2018)].

(3) Evaluation:

According to a published text, evaluation metrics, including accuracy, precision, and recall, are the best for gauging how well an ML algorithm can classify a given task [Bourke, (2019)]. Contextually, for detecting or classifying a phishing website, other evaluation metrics, such as false negative rate and performance speed are considered [Basit et al., (2020)]. For instance, in a false negative scenario, a website is presumed benign when it is harmful and can further hurt business [Joshi et al., (2023)]. Also, other ML evaluation metrics for success and their calculations can be reviewed in other published texts [Bahaghighat et al, (2023)], [Hota et al., (2018)].

(4) Features:

Features can simply be defined as forms of data within a dataset [Bourke, (2019)]. In the case of website data, features are taken from the page source codes and Uniform Resource Locators (URLs) which serve as identifiers and links for specific web addresses [Calzarossa *et al*., (2023)]. Subsequently, these qualities must consider both the traits that set the two groups of websites (phishing and non-phishing) apart as well as the tactics used by attackers to trick people. When it comes to machine learning, phishing detection methods differ mostly in the attributes they use to characterize the characteristics of the phishing vectors and the learning algorithms they use to categorize them [Hota et al., (2018)]. Some publications' features (i.e., URL-based features) pertain to the phrasal and numerical attributes of URL threads [Rao and Pai, (2019)], while other papers' features (i.e., HTML-based features) refer to the information about a page's content and aesthetics [Corona et al., (2017)].

The authors in [Calzarossa *et al*., (2023)] argue that because they don't involve downloading any pages, URL-based characteristics are typically quick to extract. This makes it possible to detect phishing websites in real-time and defend against novel phishing attacks, popularly named zero-hour strikes. However, these attributes are vulnerable to manipulation by attackers who frequently employ link manipulation techniques to trick unsuspecting victims into thinking that a link comes from a reliable and trustworthy source. Conversely, HTML-based functionality exhibits resilience against evasion tactics, mystification, and disguising techniques commonly employed by phishers. However, the removal of these functionalities could cause delays in the download of the website as well as safety and security concerns. Also, the introduction of extraneous applications, including web search tools, can be used to extract features but can introduce delays [Rao and Pais, (2019)].

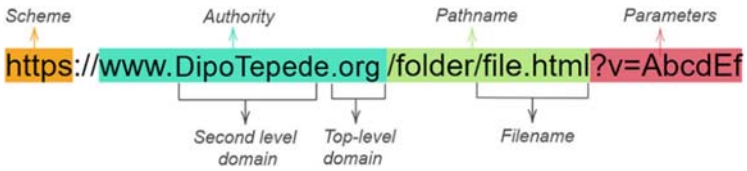

Figure 2: Structure and components of a URL.

As illustrated in Figure 2, the major components of a URL [Stallings, (2020)] are,

Scheme: This tells you what protocol to use to access the resource.

Authority: This attribute provides information regarding the internet location of the resource while simultaneously identifying the associated Domain Name.

Path: This tells you the specific resource that you want to access on the website.

Query: This is an optional part of the URL that can be used to pass parameters to the resource.

In general, the most important characteristics for distinguishing phishing websites from non-phishing ones are the URL and pathname length [Calzarossa *et al*., (2023)].

(5) Modeling:

In this framework [Bourke, (2019)], modeling can be broken into three stages: selecting a prototype, tuning a prototype, and comparing it with others.

(i) Selecting a prototype

Performance efficiency is an important consideration when choosing an ML-based phishing solution [Balogun *et al*., (2021)]. For instance, in making a final decision to eradicate blockchain phishing attempts, the authors in [Joshi et al., (2023)] selected XGB over RF with a higher accuracy rate, due to XG Boost's lower False Negative Rate and performance Speed (see Table 1).

ML Model	Accuracy Rate	False Negative Rate	Performance Speed
Random Forest	98.50%	1.83%	3.01 s
XG Boost	98.33%	1.77%	2.03 s.

Table 1: Selecting Model by Comparing Evaluation Metrics [Joshi et al., (2023)]

Nevertheless, in another study [Calzarossa *et al*., (2023)], these ensembles of tree models (RF and XGB), well-known for their top performance in classification applications [Mitchell, (1997)], may fail in "explainability", which is the capacity to evaluate the feature's relative value or influence to the whole ML model successful performance. Consequently, the study [Calzarossa *et al*., (2023)] closes that gap by advocating a method, underpinned by the Gini index, which shows explainable variables, employing a univariate exploratory examination. Based on the RF method and this feature selection method, the authors of the study developed a multidimensional predictive model for phishing.

Also, ML models can be selected based on the form of the dataset [Bourke, (2019)]; Deep Learning models such as neural networks broadly align with unstructured data including images and audio files compared to Decision Tree derivatives and ensembles that align best with structured data. However, the trade-off is that Deep Learning models usually use longer training time, are difficult to debug, and take more extended prediction time.

(ii) Tuning a prototype

The authors in [Hota et al., (2018)] claim that using the Remove-Replace Feature Selection Technique (RRFST) with an ensemble of two Decision Trees (C4.5 and Classification and Regression Tree (CART)) with only 11 features were able to improve their detection performance of phishing E-mail to 99.27% accuracy. However, since RRFST is a finite-state transducer (FST), it is important to highlight its drawbacks which include being computationally expensive to train and use, being difficult to scale to large datasets, and being sensitive to the order of the features [Zhang and Wang, (2023)].

Also, while some studies [Calzarossa *et al*., (2023)], [Zheng and Jin, (2018)], [Alani and Tawfik, (2022)] record effective results in feature selection methods to improve performance in ML, authors in [Bahaghighat et al, (2023)] downplay its importance and success in phishing detection, claiming that using Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) was not capable of improving the model's overall accuracy but reduce the runtime from 1.5 seconds to an insignificant 0.86 seconds.

Additionally, as presented in [Orunsolu et al, (2022)], the authors achieved exceptional results of 99.96% accuracy and a 0.04 false-positive rate through 10-fold cross-validation using Support Vector Machine (SVM) and Naïve Bayes (NB) models trained on an attribute set consisting of 15 dimensions. Although the authors may have implied that the cross-validation technique enhanced the performance of this robust system, other techniques such as using near-balanced phishing and non-phishing datasets and making use of feature selection methods may have played a significant role in such performance.

(iii) Comparing Models

Authors in [Bourke, (2019)] suggest that when comparing models, we should ensure we compare their performance based on similar contexts of operation. For instance, in [Joshi et al., (2023)] as illustrated in Table 1, XG Boost was chosen over RF but in [Calzarossa *et al*., (2023)], RF was chosen over XG Boost. Hence, beyond the performance rate of the algorithm, there exist other considerations and factors, such as robustness, interpretability, cost, and social and ethical implications, that should be considered when comparing models in real-world applications [Chen et al., (2023)]. For example, the robustness of the anti-phishing algorithm was a crucial consideration in [Orunsolu et al, (2022)] when comparing their proposed approach to related works. Also, in [Calzarossa *et al*., (2023)], the interpretability of the ML anti-phishing solution was a vital decision in its study.

(6) Experimentation:

The focus of this step is to reduce the time between offline investigations and online investigations [Bourke, (2019)]. To achieve this, the exchange of training, validation, and test data is not a topic that the author of [Mitchell, (1997] specifically covers in his work. However, the author discusses cross-validation without explicitly mentioning the idea of sharing data. Cross-validation is a resampling technique for assessing the generalizability of a model by iteratively partitioning the dataset into training and validation sets, thereby simulating performance on unseen data [Orunsolu et al, (2022)].

The initial step in cross-validation involves partitioning the data into k groups. Subsequently, the model undergoes training on k - 1 groups while the unused group is used for evaluation, iteratively across all folds. The model's performance on unobserved data is then estimated using the average performance across all k folds. Notwithstanding, the majority of publications merge training and validation dataset proportions and concur on a split in the following proportion range; Training dataset (70–80%), Validation/development dataset (10–15%), and Test dataset (10–15%) [Alani and Tawfik, (2022)], [Orunsolu et al, (2022)], [Bahaghighat et al, (2023)], [Joshi et al., (2023)], [Ahammad et al., (2022)].

3.0 Methodology

An extensive literature review was carried out by searching various online databases, such as Research Gate and Science Direct, for journal articles associated with ML Model solutions for Detecting and Classifying Websites. Also, our focus primarily rested on the ACM digital library and the Springer database, recognized as comprehensive data sources within the computing domain [Valente et al., (2022)]. To guarantee the utilization of relevant and recent articles, the search was limited to articles published between 2018 and 2023.

Following meticulous examination and analysis of pertinent literature on website detection and classification through ML models, identified gaps in the existing state of solutions were considered during the model design phase. The central aim of this work is to outline the sequential modeling steps aimed at countering phishing attacks, with a summary of findings from ten relevant papers presented in Table 2.

Source Articles	Problem	Data	Evaluation	Features	Modelling	
					Selecting	Tuning
Calzarossa1 et al (2023)	Binary Classification	Streaming Data	O.96 AUC	6 Feature Types.	Random Forest	Gini Index
Joshi et al (2023)	Binary Classification	Streaming Data	98.5% & 98.33 accuracy respectfully	15 Features	Random Forest & XG Boost	Gini Index
Bahaghighat et al (2023)	Detection	Streaming and Static Data	99.2% accuracy, 99.1% precision, 99.4% recall, and 99.1% specificity	111 Features	XG Boost	Principal Component Analysis (PCA)
Ahammad et al (2022)	Detection	Streaming Data	0.895 accuracy	15 features: classified	LightGBM	None

Orunsolu et al (2022)	Detection	Streaming Data	99.6% & 99.6% accuracy respectfully	15-dimensional feature	SVM & NB	Cross-Validation
Mahmoud & Hesham (2022)	Detection	Preprocessed datasets	97.49% & 98.69% accuracy respectful for both data set	111 features into 6 group	RF, KNN, DT, LDA and BNB	None
Alani and Tawfik (2022)	Detection	Static URL	97.5%. accuracy	14 Features	Random Forest	Recursive feature elimination
Balogun et al (2021)	Detection	3 Static Dataset	98.51% accuracy	30, 48, 10 Features respectively	Functional Tree (FT) and its variants. BT-FT-2	Rotational Forest Meta Learners
Chin et al (2018)	Binary Classification	Static Data	98.39% Accuracy	30-dimension vector and Domain Features	ANN	Partial Least Squares
Hota et al (2018)	Detection	Static Data	99.27% Accuracy	11 Features	Ensemble of DT (C4.5 and CART)	Remove-Replace Feature Selection Technique

Table 2: 10 Recent Scholarly Study on Phishing Prevention using the framework in [Bourke, (2019)]

4.0 Findings

Websites are common phishing vectors for numerous psychological manipulation assaults on the internet, including numerous continuing web frauds [Shaikh, (2016)] [Naqvi, *et al*, (2023)]. Assailants in such attacks mimic legitimate website pages and distribute suspected URLs to their intended casualties via unsolicited emails, text messages, or other internet media platforms. These attackers generally disseminate a fake version of a real website by email, phone, or text messaging [Basit *et al*., (2020)], hoping that the recipients will fall for the hoax contained in the email content. Their ultimate purpose is to compel victims into providing personal or extremely confidential data, such as debit cards and web login information. Consequently, the opportunity for FBA phishing commences with any of the IB users inputting their sensitive information into a duplicate website [Parekh *et al*., (2018)], [Shaikh, (2016)], [Naqvi, *et al*, (2023)].

The ML modeling steps start by framing the business problem into an ML problem. The primary business problem is to enhance the security of Internet banking services in Nigeria by effectively detecting and preventing Fake Bank Alert phishing attacks. These attacks involve tricking customers into believing they have received legitimate bank alerts when, in fact, they are fraudulent [Kagantech (2023)]. However, without the sensitive information stolen from the unsuspecting victim, the fraud cannot be successful [Michael, (2022)], hence, implementing supervised ML systems directly on each endpoint enables the real-time detection of threats, functioning within seconds even when a device operates offline. [Columbus, (2020)].

Data can be collected from historical fraud incident data and phishing and non-phishing website repositories. The historical fraud incident anonymized data can be collected from selected Banks in Nigeria especially those with significant principal brands and investments in IB capacities, as they are highly ranked for their excellent customer-centric philosophy [KPMG, (2014)]. This selection ensures the authenticity of the experiment results [AbuShanab, (2010)]. Also, other avenues to source active phishing websites include PhishTank, a repository platform that collects, verifies, shares, and monitors phishing data. Other repositories, such as Tranco (known for ranking millions of domains), can source legitimate websites [Calzarossa *et al*., (2023)].

Decision Tree ensembles, such as XG Boost and Random Forest, were selected as our models of choice. The study highlighted the promising outcomes of these models regarding accuracy, precision, and mitigation of false positives and negatives [Calzarossa *et al*., (2023)] [Joshi et al., (2023)]. Specifically, the selection of features, encompassing URL-based attributes like URL length and pathname, demonstrated efficacy in discerning between

phishing and legitimate websites [Calzarossa *et al*., (2023)]. Furthermore, the research underscored the role that feature selection plays in improving the model's overall performance.

After deploying the model, continuous monitoring and testing are essential to ensure it behaves as expected and adapts to evolving Fake Bank Alert phishing tactics. The proposed model's steps may need to be revisited iteratively based on the model's performance in real-world scenarios, and adjustments should be made accordingly. Other avenues for exploration include incorporating natural language processing (NLP) techniques to analyze alert message content.

4.1 The Suggested Theoretical Model

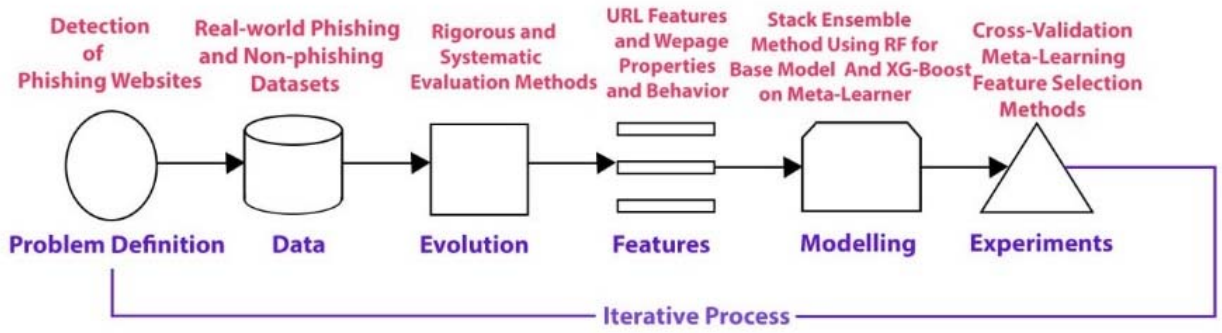

Figure 3: The Suggested Theoretical Model to counter FBA Phishing Fraud

(1) Problem Definition: This can be framed as a phishing website detection challenge since the effectiveness of this fraud requires a fake website to collect information.
(2) Data: The data will likely be structured streaming (real-time) labeled and non-phishing URLs.
(3) Evaluation: Success can be evaluated based on 95% accuracy, 95% precision, 95% recall, 5% False Positive, and 5% False Negative
(4) Features: Selected features will be based on URL Length, URL Domain Name, URL pathname, URL filename, and URL parameters. Webpage properties and behaviors may be added.
(5) Modeling:
    (i) Model Selection: Stack Decision Tree Ensembles using XG Boost or Random Forest.
    (ii) Model Improvement: Model performance can be improved through hyper-parameter tuning, feature selection, and cross-validation.
    (iii) Comparison: Model performance should be compared with baseline models in a similar context to identify the most effective approach.
(6) Experimentation: Data split will be based on Cross-Validation: Training dataset (90%), and Test dataset (10%)

5.0 Conclusion:

This study presents a comprehensive approach to address the pressing challenge of Fake Bank Alert (FBA) phishing attacks within Nigeria's Internet Banking system. Grounded in an extensive literature review and meticulous analysis, the research underscores the sophistication of these attacks, highlighting the need for a robust defense mechanism.

The findings reveal the pervasive nature of phishing as a gateway to various online social engineering attacks, emphasizing the replication of legitimate websites to extract sensitive information. Leveraging machine learning (ML), the proposed model focuses on enhancing security by swiftly detecting and thwarting FBA phishing attempts.

The structured ML framework, rooted in problem definition, data collection sources, model selection (favoring Decision Tree Ensembles), and feature significance, offers a roadmap for a potent defense system against evolving

phishing tactics. Continuous model monitoring, testing, and potential integration of natural language processing (NLP) techniques stand as imperative steps in fortifying defenses against emerging threats.

The proposed conceptual model encapsulates crucial stages—from framing the problem to data collection, model selection, and improvement—culminating in a structured approach aimed at achieving high accuracy and precision in phishing website detection.

This research advocates a holistic ML-driven defense strategy, amalgamating technological advancements with continual refinement, to fortify Nigeria's Internet banking landscape against the persistent threat of FBA phishing attacks.

5.1 Future Recommendation

To enhance the generalizability of the proposed ML model, expanding dataset diversity is crucial in refining the accuracy and adaptability of our ML model, facilitating better differentiation between genuine and counterfeit websites. Also, ensuring training and testing are done in real-time since crafting a system for on-the-fly website detection and classification eliminates reliance on user reports, ensuring a more proactive defense against fraudulent sites. Furthermore, a performance evaluation grounded in real-world data is imperative. It's the litmus test, providing a pragmatic gauge of our model's actual effectiveness in shielding against FBA phishing threats.

Increasing awareness about FBA phishing attacks is pivotal for enhancing the efficacy of ML models in detecting phishing attacks within the Nigerian context. Banks should prioritize investments in robust security awareness campaigns aimed at educating customers about the perils associated with phishing attacks and empowering them to identify suspicious activities. Informed users can act as an additional layer of defense. Also, financial institutions in Nigeria should consider collaborating and sharing anonymized data related to fraudulent activities with researchers and scholars focused on contributing to this aspect of cybersecurity knowledge.

There is also the role of government to enact adequate laws and policies concerning cybersecurity, for instance, the Nigerian regulatory bodies should consider developing and implementing a robust regulatory framework specifically addressing IB security. This framework could mandate security standards and regular audits to ensure that banks effectively protect their customers from phishing threats. Furthermore, it is essential to address ethical and privacy concerns. Future research should explore the ethical implications of using ML for security and develop guidelines for responsible implementation.

As mentioned earlier in the paper, exploring NLP techniques to analyze the content of bank alert messages could provide valuable insights into detecting fraudulent messages.

Conflict of Interest

The authors have no conflicts of interest to declare. The authors have seen and agree with the contents of the manuscript and there is no financial interest to report. We certify that the submission is original work and is not under review at any other publication.

6.0 References

[1] AbuShanab E., Pearson J. M., and Setterstrom A. J. (2010): Internet Banking and Customers‟ Acceptance in Jordan: The Unified Model‟s Perspective, *Communications of the Association for Information Systems*, Vol. 26 No. 1, p. 494-524.

[2] Ahammad S.K.H., *et al.* (2022): “Phishing URL detection using machine learning methods” *Advances in Engineering Software* 173 103288, Available: www.elsevier.com/locate/advengsoft [Accessed September 7, 2023].

[3] Alani, M. M. and Tawfik, H. (2022): PhishNot: A Cloud-Based Machine-Learning Approach to Phishing URL Detection, *Computer Networks*, Vol. 218, https://doi.org/10.1016/j.comnet.2022.109407. Available: https://www.sciencedirect.com/science/article/pii/S1389128622004418) [Accessed September 7, 2023]

[4] Al-Qahtani, A. F., & Cresci, S. (2022, July 4): The COVID-19 scamdemic: A survey of phishing attacks and their countermeasures during COVID-19. *IET Information Security*, 16(5), 324–345. https://doi.org/10.1049/ise2.12073.

[5] Alsayed A. and Bilgrami A. (2017) “E-banking Security: Internet Hacking, Phishing Attacks, Analysis and Prevention of Fraudulent Activities”, *Int. J. Of Emerg. Techn. And Adv.* Activ, Vol.7, No.1, p.109-115.

[6] Amini M. R., Sadeghian A., & Aghaie S. (2022): Random forests outperform deep neural networks for phishing detection, *Information Sciences*, 575, 123-136.

[7] Arachchilage, N. A. G. and Love, S. (2014) Security Awareness of Computer Users: A Phishing Threat Avoidance Perspective. *Computers in Human Behavior*, Vol. 38, p.304-312.

[8] Arachchilage, N. A. G., Love S., and Beznosov, K. (2016) Phishing Threat Avoidance Behavior: An Empirical Investigation, *Computers in Human Behavior*, Vol. 60, p.185-197.

[9] Aribake F. O. and Zahurin M. A. (2020), Modelling the Phishing Avoidance Behavior Among Internet Banking Users in Nigeria: The Initial Investigation, *Journal of Computer Engineering and Technology (JCET),* Vol. 4, Iss. 01, p.1-17.

[10] Bahaghighat M., Ghasemi M., and Ozen F. (2023): A high-accuracy phishing website detection method based on machine learning, *Journal of Information Security and Applications*, Vol.77. https://doi.org/10.1016/j.jisa.2023.103553. Available: https://www.sciencedirect.com/science/article/pii/S2214212623001370 [Accessed September 7, 2023]
[11] Balogun A.O., et al. (2021), Improving the phishing website detection using empirical analysis of Function Tree and its variants. *Heliyon, Vol.* 7, No.7, June 2021. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8264617/ [Accessed September 7, 2023].
[12] Basit, A., *et al.* (2020, October 23). A comprehensive survey of AI-enabled phishing attacks detection techniques. Telecommunication Systems, 76(1), 139–154. https://doi.org/10.1007/s11235-020-00733-2
[13] Bharat, K., & Broder, A. (1998): A technique for measuring the relative size and overlap of public Web search engines. *Computer Networks*, 30(1–7), 107–117.
[14] Bourke D. (2019): A 6 Step Field Guide for Building Machine Learning Projects. [Online] Available: https://www.mrdbourke.com/a-6-step-field-guide-for-building-machine-learning-projects/ [Accessed September 7, 2023].
[15] Calzarossa M. C., Giudici P., and Zieni R. (2023): Explainable machine learning for phishing feature detection. Wiley. Available: DOI: 10.1002/qre.3411[Accessed September 7, 2023]
[16] Chen Y., Zhang Y., and Liu K., (2023): Factors affecting the success of machine learning models in real-world applications, *Machine Learning*, Vol.109, No.1, p.5-23.
[17] Chiew K.L., et al. (2019): A new hybrid ensemble feature selection framework for machine learning-based phishing detection system, *Inf. Sci*. 484 (2019) 153–166.
[18] Columbus, L. (2020): 5 Ways ML Can Thwart Phishing Attacks. Forbes [Online]. Available: https://www.forbes.com/sites/louiscolumbus/2020/08/12/5-ways-machine-learning-can-thwart-phishing-attacks/?sh=6b8698871035 [Accessed September 7, 2023].
[19] Corona I., Biggio B., and Contini M. (2017): DeltaPhish: Detecting phishing webpages in compromised websites. In: Foley S, Gollmann D, Snekkenes E, eds, *Computer Security —ESORIC. Lecture Notes in Computer Science.* Vol.10492. Springer; p.370-388, 2017.
[20] Fujita M., *et al*. (2015): An Attempt to Memorize Strong Passwords while Playing Games. In 2015 18th International Conference on Network-Based Information Systems (NBiS). 2015, pp. 264-268.
[21] Gomes, L., Deshmukh, A., & Anute, N. (2022, July 14): Cyber Security and Internet Banking: Issues and Preventive Measures. *Journal of Information Technology and Sciences*, 8(2), 31–42. https://doi.org/10.46610/joits.2022.v08i02.005
[22] Gupta B.B., Arachchilage N.A., and Psannis K.E. (2018): Defending against phishing attacks: taxonomy of methods, current issues and future directions, *Telecommun.* Syst. 67 (2), pp. 247267.
[23] He, D., Chan, S., & Guizani, M. (2015, February). Mobile application security: malware threats and defenses. *IEEE Wireless Communications*, 22(1), 138–144. https://doi.org/10.1109/mwc.2015.7054729
[24] Hota H.S., Shrivas A.K., Hota R., (2018): An Ensemble Model for Detecting Phishing Attack with Proposed Remove-Replace Feature Selection Technique, *Procedia Computer Science*, Vol. 132, p.900-907, 2018. https://doi.org/10.1016/j.procs.2018.05.103. Available: https://www.sciencedirect.com/science/article/pii/S1877050918308354
[25] Jamil H., Hussain F., and Khan S. A. (2021): Phishing Attacks: A Review, *IEEE Communications Surveys & Tutorials,* Vol.23, No. 4, pp.2835-2869.
[26] Jansen J. (2015), Studying Safe Online Banking Behavior: A Protection Motivation Theory Approach. *In HAISA*, pp. 120-130. 2015.
[27] Joshi K., et al. (2023): Machine-Learning Techniques for Predicting Phishing Attacks in Blockchain Networks: A Comparative Study, *Algorithms*, Vol. 16, No. 366. 2023. Available: https://doi.org/10.3390/a16080366 [Accessed September 7, 2023]
[28] Kagantech (2023): Exposing the Underbelly of Fake Transfer Fraud!!. Instagram [Online]. Available: https://www.instagram.com/p/Cv6f5C_IdDe/ [Accessed September 7, 2023].
[29] Kanhere S., *et al.* (2018): Phishing: A Comprehensive Study, *IEEE Communications Surveys & Tutorials,* Vol.20, No.4, p.2756-2789.
[30] Khedmatgozar, H. R., & Shahnazi, A. (2017, March 8): The role of dimensions of perceived risk in adoption of corporate internet banking by customers in Iran. *Electronic Commerce Research*, 18(2), 389–412. https://doi.org/10.1007/s10660-017-9253-z
[31] KPMG (2014): Banking industry customer satisfaction survey. KPMG Professional Services, Lagos, available at: www.Kpmg.com.
[32] Lokesh G. H. and BoreGowda, G. (2020) Phishing website detection based on effective machine learning approach, *J. Cyber Sec. Technol*. 114.
[33] Mahmoud O. and Hesham H. (2022): An Empirical Study Towards an Automatic Phishing Attack Detection Using Ensemble Stacking Model. *Future Computing and Informatics Journal*: Vol. 7: Iss.1, Article 1. DOI: http://Doi.org/10.54623/fue.fcij.7.1.1 Available at: https://digitalcommons.aaru.edu.jo/fcij/vol7/iss1/1 [Accessed September 7, 2023]
[34] Michael, C. (2022): Fake bank alerts put small businesses at risk. BusinessDay [Online]. Available: https://businessday.ng/technology/article/fake-bank-alerts-put-small-businesses-at-risk/ [Accessed September 7, 2023].
[35] Mitchell T. (1997): Machine Learning. McGraw-Hill International Edition.
[36] Mitrobe (2022): Top 5 Apps in Nigeria That Does Fake Bank Alerts. NGSUP NETWORK [Online]. Available: https://ngsup.com/top-5-apps-in-nigeria-that-does-fake-bank-alerts/#ixzz7ZHcohx4O [Accessed September 7, 2023].
[37] Mitrobe Network (2022): Scam Alert| 5 Apps That Can Be Used to Do Fake Bank Alerts. YouTube [Online]. Available: https://www.youtube.com/watch?v=poR9GbKhIG0 [Accessed September 7, 2023].
[38] Mridha F. M., Nur K., Saha K. A., & Adnan A. (2017, February 15). A New Approach to Enhance Internet Banking Security. *International Journal of Computer Applications*, 160(8), 35–39. https://doi.org/10.5120/ijca2017913093
[39] Naqvi, B., et al (2023), "Mitigation strategies against the phishing attacks: A systematic literature review", *Computers & Security*, Vol.132. Elsevier Ltd. Available: https://doi.org/10.1016/j.cose.2023.103387. [Accessed September 7, 2023]
[40] Orunsolu A.A., Sodiya A.S., and Akinwale A.T. (2022): A predictive model for phishing detection, *Journal of King Saud University – Computer and Information Sciences*, Vol. 34, p. 232–247. Available: https://doi.org/10.1016/j.jksuci.2019.12.005 [Accessed September 7, 2023]
[41] Othman M. and Hassan H. (2022): An Empirical Study Towards an Automatic Phishing Attack Detection Using Ensemble Stacking Model. *Future Computing and Informatics Journal*, Vol. 7, Iss 1.
[42] Parekh S., *et al.* (2018): A new method for detection of phishing websites: Url detection. *In 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT),* p.949–952, 2018. IEEE.
[43] Rahman S.S.M.M, et al. (2020): Performance assessment of multiple machine learning classifiers for detecting the phishing URLs, in: *Data Engineering and Communication Technology*, Springer, pp. 285–296.
[44] Rao R. and Pais A. (2019): Detection of phishing websites using an efficient feature-based machine learning framework, *Neural Comput Appl*., Vol. 31, No. 8, p.3851–3873.
[45] Shaikh A., Shabut A., and Hossain A. (2016), A literature review on phishing crime, prevention review and investigation of gaps, p.9-15. Available: https://ieeexplore.ieee.org/document/7916190 [Accessed September 7, 2023].
[46] Shie E. W. S. (2020): Critical analysis of current research aimed at improving detection of phishing attacks, Selected computing research papers, p. 45. 2020.

[47] Stallings W. (2020): Data and computer communications, (11th ed.). Pearson.
[48] Urias, V.E., et al. (2017), Technologies to enable cyber deception, in: 2017 International Carnahan Conference on Security Technology (ICCST), *IEEE*, pp. 16.
[49] Valente, A., et al. (2022): Analysis of Academic Databases for Literature Review in the Computer Science Education Field. In: 2022 IEEE Frontiers in Education Conference (FIE), IEEE, pp. 1–7.
[50] Volkamer M., et al., (2017): User experiences of TORPEDO: TOoltip-poweRed Phishing Email DetectiOn", *Computers & Security*, Vol.71, p.100-113, 2017. ISSN 0167-4048, Available: https://doi.org/10.1016/j.cose.2017.02.004. [Accessed September 7, 2023].
[51] Vrbanˇciˇc G., Fister Jr I., & Podgorelec V. (2018): Swarm intelligence approaches for parameter setting of deep learning neural network: Case study on phishing websites classification, In Proceedings of the 8th international conference on web intelligence, mining and semantics, p.1–8,.
[52] Yang P., Zhao G., and Zeng P. (2019): Phishing website detection based on multidimensional features driven by deep learning. *IEEE Access 7*, p.15196–15209.
[53] Zhang X. and Wang J. (2023): A comparative study of feature selection techniques for finite-state transducers, *Information Sciences*, vol. 575, p.123-136.
[54] Zheng W. and Jin M. (2018) A Comparative Study Of Feature Selection methods", International Journal on Natural Language Computing (IJNLC), Vol.7, No.5. Available: https://www.researchgate.net/publication/328904014_A_Comparative_Study_of_Feature_Selection_Methods [Accessed Sep 15 2023].
[55] Zhu E., et al (2020): DTOF-ANN: an artificial neural network phishing detection model based on decision tree and optimal features. *Appl. Soft Comput*., 106505.

Authors Profile

Dr. Seun Ebiesuwa holds a first degree in Computer Engineering from the Lagos State University. Dr. Ebiesuwa bagged a Master's and a Ph.D degree in Computer Science from Babcock University. He is a lecturer in the Department of Computer Science, Babcock University and he specializes in Information Systems. Dr. Seun Ebiesuwa has over forty-five academic publications in internationally peer-reviewed journals of Computer Science in the areas of Artificial Intelligence, Data Analytics, Machine Learning and Medical Informatics. He has made several academic presentations in Workshops, Seminars, Symposiums and Conferences.

Dipo Tepede is a recognized expert in Process and Project Management, holding a Master Black Belt in Lean Six Sigma, along with certifications as a Professional Project Manager and Business Analyst from the Project Management Institute, and as a SAFe Scrum Master from Scaled Agile. He obtained his undergraduate degree from Obafemi Awolowo University and an MBA from the University of Gavle, Sweden. His academic journey expanded to Arden University in the United Kingdom, where he explored a Master of Science in Data Analysis. Currently pursuing a Ph.D. in Cybersecurity at Babcock University, his research focuses on Machine Learning, Data Analytics, and Artificial Intelligence, showcasing his commitment to advancing knowledge in the field.