International Journal of Modern Developments in Engineering and Science
Volume 3, Issue 3, March 2024
https://www.ijmdes.com | ISSN (Online): 2583-3138

Email Anti-Phishing: Machine Learning Models and Evaluation Overview

Obianuju Nwaogo Mbadiwe[1*], Obi Chukwuemeka Nwokonkwo[2], Charles O. Ikerionwu[3], Anthony Ifeanyi Otuonye[4], Chukwuemeka Etus[5], Christiana Amaka Okoloegbo[6]
[1,2,3,4,5,6]*Department of Information Technology, School of ICT, Federal University of Technology, Owerri, Nigeria*

***Abstract*: Phishing attacks have grown to be one of the most visible and challenging issues confronting internet users, organizations, and governments. To effectively combat phishing attacks, it is imperative to have robust machine learning models for email anti-phishing systems. These models play a crucial role in analyzing email content, sender behavior, and other relevant features to identify and block potential phishing emails. To make sure these machine learning models work well in real-world scenarios, it is crucial to evaluate their performance. This paper has reviewed machine learning anti-phishing solutions through a systematic literature review considering the integration of diverse machine learning techniques, including ensemble models, coupled with advanced evaluation methodologies. This review concludes that Email security has improved significantly with the application of machine learning to counter phishing attempts. Also, the incorporation of machine learning models into anti-phishing tactics has resulted in the creation of resilient defenses against the ever-growing sophistication of cyber threats.**

***Keywords*: Accuracy, Cyber threats, Email, Evaluation, Machine learning, Phishing.**

1. Introduction

One of the most common and difficult problems that affect internet users, businesses, and governments nowadays is phishing attacks [1]. Through phony emails, cybercriminals deceive victims into divulging private information like passwords, credit card numbers, or social security numbers. While there are technical anti-phishing solutions that use AI-powered and techniques based on machine learning to spot and prevent phishing emails, these solutions cannot stop all emails with phishing content. The over 700 percent increase in number of phishing sites between the 3rd quarter of 2020 and last quarter of 2022 highlights the growing sophistication of cybercriminals and the challenges faced by traditional anti-phishing solutions [2]. The increased reliance on internet applications has provided cyber attackers with numerous opportunities to steal and misuse user data, making it imperative to explore advanced methods for countering phishing threats. In order to effectively counter phishing attacks, the development of robust machine learning models is essential for email anti-phishing systems. These models are pivotal in scrutinizing email content, sender behavior, and other pertinent attributes to detect and prevent potential phishing emails. Ensuring the efficacy of these machine learning models in real-world scenarios necessitates thorough performance evaluation.

Moreover, the significance of employing advanced evaluation methods to assess the efficacy of email anti-phishing models cannot be overstated. Cross-validation, a commonly used evaluation technique, allows for the rigorous testing of models by systematically splitting the dataset into multiple subsets. However, the exploration of novel evaluation methodologies that can comprehensively assess the performance and robustness of anti-phishing models is crucial for further advancements in this domain [3]. Given how frequently phishing attacks occur and how cyber security threats have evolved, the quest for effective anti-phishing solutions remains a paramount concern. It is imperative to continue pushing the boundaries of innovation, leveraging cutting-edge technologies, and evolving evaluation methodologies to bolster the resilience of email anti-phishing systems against the ever-growing sophistication of cyber threats [4].

2. Background and Related Literature

Subsection 2A provides background on machine learning models used for email anti-phishing and evaluation. Subsection 2B follows with a review of related work.

A. Background on Machine Learning Models and Evaluation Metrics Used for Email Anti-Phishing

1) Machine Learning Models for Email Anti-Phishing

In recent times, researchers have been developing predictive mechanisms and algorithms using deep learning and machine learning methods to combat phishing attacks. These efforts have led to the proposal of innovative solutions such as the THEMIS deep learning algorithm, which focuses on modeling email bodies and headers at the character and word levels, using Regional-based Convolutional Neural Network (R-CNN) for phishing email prediction. THEMIS stands out as a cutting-edge deep learning (DL) model designed to identify phishing emails. Survey findings reveal that THEMIS, which uses multilevel vectors and recurrent neural networks, has an excellent accuracy rate of 99.848%. It is worth noting that the model's sole limitation lies in its inability to identify phishing

*Corresponding author: mbadiweobianuju.20174076878@futo.edu.ng
Date of Publication: 14-08-2024
DOI: 10.5281/zenodo.13327914

attempts in emails lacking an email header, despite its overall effectiveness in detecting such threats [5]. Furthermore, the alignment of phishing emails with a user's work context poses a significant challenge in detection. Artificial intelligence techniques, including machine learning, deep learning, hybrid learning, and scenario-based techniques, are being urgently explored to address this issue. The urgency in exploring these advanced techniques underscores the critical nature of the phishing email pandemic.

In addition to advanced AI techniques, the study of phishing email detection has explored approaches such as Deep Learning, blacklisting, and machine learning-based classification algorithms. While blacklisting methods depend largely on personal reports, phishing emails continue to evade identification despite significant time and workforce allocation. These challenges emphasize the need for more sophisticated and effective anti-phishing strategies, calling for continued research and innovation in this critical cybersecurity domain. Ensemble models, which combine multiple machine learning models, have also shown promise in email anti-phishing. Ensemble models combine the predictions of multiple individual models to make a final prediction. According to Kim Soon et al. [6] ensemble methods are known to achieve better accuracy in phishing prediction. The inadequacy of current techniques in effectively countering phishing attacks underscores the dire need for more advanced and innovative anti-phishing solutions. The increasing sophistication of cybercriminals, coupled with the rising number of phishing attacks, demands a paradigm shift in the approach to email anti-phishing [3]. In light of these challenges, the integration of ensemble models, which combine multiple machine learning models, has shown promise in enhancing the predictive capabilities of email anti-phishing systems. The utilization of ensemble models holds great potential in improving the overall accuracy and reliability of phishing email detection, thereby offering a more robust defense against these malicious attacks [7].

2) Understanding Machine Learning Models

The purpose of machine learning models is to use training data to identify patterns and generate predictions [8]. These models can identify incoming emails as phishing or legitimate based on features extracted from the email text, sender information, and other pertinent qualities by training them using a labeled dataset of phishing and legitimate emails [9]. Machine learning models are algorithms that, without explicit programming, use data to learn and generate predictions or assessments [10]. These models are trained using labeled datasets, in which every email is categorized as either legitimate or phishing [11].

There are several popular machine learning models used for email anti-phishing, including:

a) *Logistic Regression:* This model is commonly used for binary classification tasks, making it suitable for distinguishing between phishing and legitimate emails [3].
b) *Decision Trees:* Decision tree models use a hierarchical structure of decision nodes to classify data based on different features [12].
c) *Random Forest:* A random forest model is an ensemble model that combines multiple decision trees to achieve higher accuracy [13].
d) *Naive Bayes:* This probabilistic model is based on Bayes' theorem and assumes that the features are independent of each other [7].
e) *Support Vector Machines:* SVM models are effective in separating data points by creating a hyperplane that maximally separates different classes [7].
f) *Neural Networks:* Neural networks are deep learning models that consist of interconnected layers of nodes, mimicking the structure and function of the human brain. These models are trained using a large amount of labeled data, and they learn to recognize complex patterns and relationships within the data [14].
g) *Gradient Boosting:* Gradient boosting models combine weak individual models (typically decision trees) sequentially, with each subsequent model trying to correct the mistakes made by the preceeding one [15].
h) *Ensemble Models:* Ensemble models combine multiple individual models to make predictions. By aggregating the predictions of multiple models, ensemble models can often achieve higher accuracy and robustness compared to individual models [16].

3) Deep Dive into Ensemble Models

Ensemble models have gained significant popularity in the field of email anti-phishing due to their ability to improve predictive accuracy and generalize well on unseen data. One of the widely used ensemble techniques is the random forest model, which aggregates the predictions of multiple decision trees to make the final classification. The diversity in decision trees within the random forest helps in reducing overfitting and capturing a wide range of patterns present in the email data [17]. Another notable ensemble technique is gradient boosting, which sequentially combines weak individual models to form a strong predictive model. By iteratively correcting the errors of the preceding models, gradient boosting can effectively learn complex relationships and improve overall prediction performance [18]. In addition to using individual machine learning models, organizations have leveraged ensemble models to enhance the robustness and reliability of their email anti-phishing systems. By integrating diverse predictions from multiple models, ensemble techniques can mitigate the limitations of individual models and provide more accurate identification and classification of phishing emails [19]. As organizations continue to combat the growing threat of phishing attacks, the utilization of ensemble models and the integration of diverse machine learning techniques will play a pivotal role in strengthening email anti-phishing defenses and safeguarding sensitive information.

4) Methods of Evaluating Anti-Phishing Models

It is imperative to take into account parameters like precision, recall, and the F1 score when evaluating the performance of ensemble models in differentiating between phishing and legitimate emails. Furthermore, continuous retraining and updating of the ensemble models with the latest email threat

data are essential to adapt to evolving phishing tactics and maintain high detection accuracy [1]. A variety of techniques, including cross-validation, which divides the dataset into different subgroups and uses each subset for training and testing data to assess the ensemble model's performance, can be used to evaluate ensemble models [20]. Holdout validation is one of the other evaluation techniques, in which the dataset is split into training and testing sets, and the ensemble model is trained on the training set and assessed on the testing set [19]. In order to gain a more reliable assessment of the ensemble model's performance, organizations can also employ methods like k-fold cross-validation, which entails splitting the data into k subsets and repeatedly training and testing the model on various combinations of subsets [19]. Evaluation measures that can be used to gauge the effectiveness of anti-phishing models include precision, recall, and the F1 score. These metrics offer information about the model's precision and recall—which measure how well it can identify and classify phishing emails—as well as its ability to balance precision, recall, and the F1 score in order to assess how well ensemble models work at correctly identifying and classifying phishing emails [21].

Organizations can also use more sophisticated evaluation methods, like area under the curve analysis and receiver operating characteristic curves, to evaluate the trade-off between true positive rate and false positive rate. This will give them a thorough understanding of the ensemble model's performance across various thresholds [22].

Furthermore, the robustness and resilience of ensemble models can be enhanced through the incorporation of feature importance analysis, which helps identify the most influential attributes in the classification of phishing emails. By understanding the relative importance of different features, organizations can refine their feature selection process and improve the overall predictive power of the ensemble model [3]. In the quest for effective anti-phishing solutions, it is essential for organizations to continuously explore and adapt novel evaluation methodologies that comprehensively assess the performance and robustness of ensemble models. Embracing a multidimensional approach to evaluation, which incorporates traditional metrics alongside advanced techniques, will empower organizations to strengthen their email anti-phishing defenses and stay ahead of the evolving landscape of cyber threats [15]. As the demand for sophisticated anti-phishing solutions continues to grow, the integration of diverse machine learning techniques, including ensemble models, coupled with advanced evaluation methodologies, will be pivotal in bolstering the resilience of email anti-phishing systems against the ever-growing sophistication of cyber threats [23].

5) Key Evaluation Metrics for Anti-Phishing Models

When evaluating anti-phishing models, there are several key metrics that are commonly used to measure their performance:

a) Accuracy: The overall accuracy of the model in correctly classifying phishing and non-phishing emails.

b) Precision: The proportion of correctly classified phishing emails out of all the emails predicted as phishing.

c) Recall: The proportion of correctly classified phishing emails out of all the actual phishing emails [19].

d) F1 score: The harmonic mean of precision and recall, providing a balanced measure of both metrics.

e) False Positive Rate: The proportion of non-phishing emails incorrectly classified as phishing.

f) AUC-ROC: The area under the receiver operating characteristic curve, which measures the performance of the model across different classification thresholds.

6) Comparative Analysis of Anti-Phishing Models

The study of cybersecurity has prominently focused on addressing phishing emails, with extensive research exploring various machine learning algorithms for their efficacy in anti-phishing models. Among these algorithms, the C5.0 algorithm stands out as a popular choice, demonstrating high precision in categorizing emails as either phishing or non-phishing [13]. Complementing C5.0, the Support Vector Machine has proven effective in accurately identifying phishing emails, leveraging its capability to classify emails within high-dimensional feature spaces.

Convolutional Neural Networks (CNNs), a particular kind of deep learning technique, have shown great promise in improving the classification accuracy of phishing emails.. CNNs excel in detecting subtle patterns and intricate relationships within email data, providing an automated feature extraction process that contributes to the effectiveness of anti-phishing solutions.

In addition to deep learning, the application of natural language processing (NLP) models, including recurrent neural networks and transformer-based models, has gained attention. These models analyze textual content, leveraging linguistic patterns and contextual cues to identify phishing attempts. NLP models excel in capturing the nuances of human language, strengthening defenses against sophisticated phishing tactics.

Anomaly detection methods, such as Isolation Forest and One-Class SVM, offer a complementary perspective by using unsupervised learning algorithms to identify anomalous behavior indicative of phishing activities. Profiling normal email behavior and flagging deviations contribute to a comprehensive anti-phishing strategy.

The dynamic nature of cybersecurity has led to the continual evolution and integration of these modern methods in anti-phishing efforts. As organizations strive to safeguard sensitive information from phishing attacks, the exploration and incorporation of diverse machine learning techniques play a crucial role in fortifying email anti-phishing defenses and effectively mitigating potential risks.

B. Related Works

Numerous investigations and literature reviews have been undertaken regarding the application of machine learning models in the context of email anti-phishing.

[24] presents a methodology to distinguish phishing emails from real ones by using a word document matrix for feature engineering and conventional machine learning techniques. Furthermore, the system incorporates lexical characteristics and domain knowledge into the feature engineering procedure. The system's efficacy is assessed by means of contrasts with diverse

traditional machine learning methodologies. For the numerical representation of phishing emails, [25] used a distributional representation, more precisely TF-IDF. They also compare and contrast traditional machine learning methods including Random Forest, AdaBoost, Naive Bayes, Decision Trees, and SVM. In terms of training accuracy, they found that Random Forest and Decision Tree performed the best. A model that separates the main dataset into n partitions according to its accessible features is presented by [26] in their paper Partitioned Dataset Cross-Fold Strategy (PDCFS). Next, every partition is trained and tested using different classifiers (Support Vector Machine, Naïve Bayes, C4.5, Random Forest, JRip, PART, and k-Nearest Neighbors) using 5-fold cross-validation. The best possible result is then determined by a majority vote averaging the outcomes. This method applies majority voting to the entire feature set as well as to smaller feature subsets that are produced through feature selection procedures, allowing for a thorough comparison between the two. Overall results show that the suggested PDCFS achieves an accuracy of 98.24%, whereas Random Forest, using a restricted feature set of 32, achieves the highest accuracy of 98.36%. In order to analyze a corpus of phishing emails, [27] proposed A.R.E.S.: Automatic Rogue Email Spotter, which uses a CNN/RNN/MLP network with Word2Vec embeddings. In order to differentiate between phishing and legitimate emails, Word2Vec plays a crucial role in capturing synaptic and semantic commonalities within the email corpus. The study aims to showcase the efficacy of word embeddings in addressing cybersecurity challenges. The outcomes highlight the potential of integrating text analytics and deep learning techniques for cybersecurity applications. In a study, [12] presented a comparative approach for phishing email detection which combines machine learning methods (Random Forest, Decision Tree, Logistic Regression, Gradient Boosting Trees, and Naive Bayes) with Natural Language Processing techniques (TFIDF, Word2Vec, and BERT). Two datasets—one balanced and one imbalanced—were included in the evaluation. Word2Vec with the Random Forest method was shown to be the best combination on the balanced dataset, whereas Word2Vec with the Logistic Regression algorithm performed better on the imbalanced dataset. Authors [28] attempted to elucidate in a different work how to extract behavior-based features as well as email content, which features are relevant for detecting Unsolicited Bulk Emails (UBEs), and which feature set is the most distinctive. The models that were presented classified UBEs with an astounding 99% overall accuracy. In a different case, [29] used deep learning methods such convolutional neural network (CNN), recurrent neural network (RNN), long short-term memory (LSTM), and multi-layer perceptron (MLP) in combination with word embedding and Neural Bag-of-ngrams for phishing email detection. They claim that word embedding with deep learning, especially LSTM, is appropriate for the anti-phishing task based on the results of their experiments. [5] offered a methodical comparison and assessment of the numerous Deep Learning (DL) and Machine Learning (ML) models that have been put forth over the last few decades for the methodical classification of phishing emails at different phases of criminal activity. The examination of the literature indicated a limited focus on phishing email detection using natural-level Natural Language Processing (NLP) techniques. Additionally, there is a need for further exploration and utilization of contemporary Deep Learning techniques in the realm of phishing email detection research. In a publication, [30] utilized data mining techniques to categorize spam emails by employing the UCI spam base dataset. They implemented Ensemble learning methods, incorporating Naïve Bayes, decision tree, ensemble boosting, and ensemble hybrid boosting classifiers. The results showed that the effectiveness of detection tasks is much enhanced by classification models based on hybrid machine learning techniques. [10] introduces MailTrout, a browser extension that uses machine learning to provide a user-friendly security solution that helps users spot phishing emails. Their study demonstrated improved usability for end users as well as high accuracy in detecting phishing emails.

3. Research Objective and Methodology

A. Objective

The aim of this study is to provide an in-depth review of machine learning-based anti-phishing solutions by systematically examining the literature. This includes the integration of various machine learning techniques, such as ensemble models, along with advanced evaluation methods.

B. Methodology

This study utilizes the systematic literature review method, which is a research process that adheres to the specific set of guidelines proposed by [31]. The review method involves creating research questions, identifying electronic resources to investigate, collecting and analyzing data, and making recommendations. This study will start by formulating research questions and listing the databases searched for email anti-phishing solutions. The process also includes searching these databases with specific keywords, applying inclusion and exclusion criteria, interpreting the results, and drawing conclusions.

1) The Research Questions

Q1: What are the current common machine learning methods used in email anti-phishing and which is the most common method?

Q2: What are the common evaluation metrics used in email anti-phishing machine learning models and which is the most commonly used?

Q3: Which feature extraction/selection techniques are most commonly used in email anti-phishing machine learning models?

Q4: Between online and manual data sources used by the researchers for emails anti-phishing models which is the most commonly used?

2) The Relevant Documents for the Review

The databases chosen to provide relevant results for this paper are selected based on specific keywords. Some of the databases examined in this review include Springer, Elsevier, ACM Digital Library, IEEE Xplore, among others.

3) Sources of Review Papers

a) Journals
b) Conference Proceedings.
c) Published Reports
d) Researchers' theses.
e) Review Articles
f) Books Chapters
g) Webpages

4) Essential keywords for the Study

The database search was conducted between January and February 2024 without any restrictions on the publication date. This search resulted in about 72 papers. A keyword analysis performed on a collection of papers downloaded from online sources and compiled in a Microsoft Word document revealed useful information, as shown in the word cloud in Fig.1 below.

Fig. 1. Word cloud for the keywords of the selected research items

5) Inclusion and exclusion criteria

The inclusion-exclusion criteria were applied at three levels, with unrelated papers being eliminated at each stage. The initial search targeted papers in the fields of computer science and engineering. However, due to the interdisciplinary nature of the term "data," articles from other fields were initially included but later excluded from the study. Only English-language papers were considered. The systematic review covered research publications from a ten-year period, from January 2014 to February 2024. Duplicate papers from multiple libraries were discarded. After the initial exclusion, 72 papers were selected, but this was later narrowed down to 36 articles based on the chosen keywords.

6) Quality evaluation

Specific criteria were used to assess the quality of the papers included in the review, and these were:

a) Papers with clearly stated objective(s).
b) Papers with a well-defined context and experimental design.
c) Papers that thoroughly document the research process.
d) Papers where the main findings are clearly presented.
e) Papers whose conclusions are directly related to the study's aims.

4. Results

This section mainly presents results of systematic review of machine learning methods of email anti-phishing solutions and responses to the research questions. Most of the answers to the research question are synthesized form Table 1.

Table 1
Systematic review of results of email phishing studies

S.No.	Citation	Machine Learning Method	Feature Extraction/Selection Method	Dataset Sourcing Method	Evaluation Metrics used
1.	[24]	Ensemble	TF-IDF	Manually Sourced	Accuracy, Precision, Recall, F1-Score
2.	[25]	Ensemble	TF-IDF	Manually Sourced	Accuracy, Precision, Recall, F1-Score
3.	[26]	Ensemble	Hybrid Method	Online Sourced	Accuracy
4.	[27]	Deep-learning	Word2Vec	Manually Sourced	Accuracy
5.	[12]	Random Forest Logistic Regresion	Word2Vec	Manually Sourced	Accuracy, Precision, Recall, F1-Score, AUC
6.	[28]	Naïve Bayes, SVM, Bagged Decision Tree, Random Forest, Extra Tree, Ada Boost, Stochastic Gradient Boosting, Ensemble	Low Variance, High Correlation, F1, mRMR, PCA,	Manually Sourced	Accuracy, Precision, Recall, F1-Score, AUC
7.	[29]	Deep-learning	Word2Vec, Neural Bag-of-Neurons	Manually Sourced	Accuracy, Precision, Recall, F1-Score
8.	[30]	Ensemble	Information Gain	Online Sourced	Accuracy, Precision, Recall, F1-Score
9.	[10]	Bidirectional Long Short-Term Memory Network (BLSM)		Online Sourced	AUC
10.	[15]	Ensemble Artificial Neural Network	Feature Importance Ranking	Online Sourced	Accuracy
11.	[32]	Naïve Bayes Logistic Regrassion	Doc2Vec TF-IDF	Manually Sourced	Accuracy, Precision, Recall, F1-Score, Balanced Detection Rate, Normalized Balanced Detection Rate
12.	[33]	SVM	Doc2Vec TF-IDF	Manually Sourced	Accuracy
13.	[18]	XGBoost		Online Sourced	Accuracy, Precision, Recall, F1-Score, specificity
14.	[34]	Random forest, Decision Tree, KNN, Logistic Regression, Naïve Bayes, SVM	TF-TDF Information gain	Manually Sourced	Accuracy
15.	[35]	SVM, Artificial Neural Network, Logistic Regression	By Manual Inspection	Manually Sourced	Accuracy
16.	[6]	Ensemble	Binary Encoding	Online Sourced	Accuracy
17.	[36]	Deep-Learning	Word Embedding	Manually Sourced	Accuracy

A. Result Discussion on Research Question 1

Fig. 2 represents the current common machine learning methods used in email Anti-phishing.

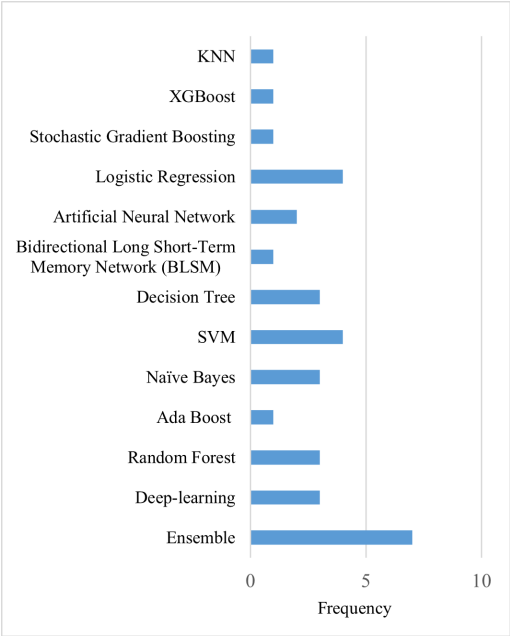

Fig. 2. Current common machine learning methods used in phishing email detection

Fig. 2 which is synthesized from Table 1 succinctly answers the research question 1, depicting the current common machine learning methods used in email Anti-phishing and it is obvious here that ensemble methods are the current prevalent methods.

B. Result Discussion on Research Question 2

To answer the research question 2, it is observed from Fig. 1 that Accuracy, Precision, Recall, F1-Score and AUC are the common evaluation metrics used in email anti-phishing machine learning models and the most commonly used metric is Accuracy.

C. Result Discussion on Research Question 3

Fig. 3 represents the common extraction/selection techniques used in email anti-phishing machine learning models.

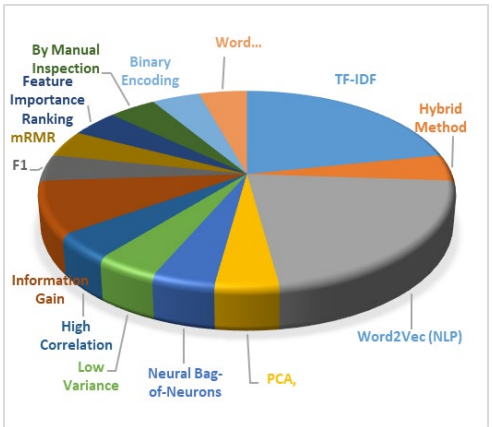

Fig. 3. Common evaluation metrics used in email anti-phishing machine learning models

Fig. 3 shows that TF-IDF and Word2Vec are the most commonly used feature extraction/selection techniques in email anti-phishing machine learning models thus answering research question 3.

D. Result Discussion on Research Question 4

In order to answer research question 4, we observe from Table 1 that manual data sources are used more by researchers for emails anti-phishing models.

5. Conclusion

Reflecting on the role of machine learning in the realm of anti-phishing, it is evident that this technology has emerged as a powerful ally in fortifying email security. The nuanced landscape of phishing attacks demands a multifaceted approach, and machine learning techniques have proven their efficacy across various dimensions. From the utilization of traditional models like Random Forest and Naïve Bayes to the integration of advanced deep learning algorithms such as convolutional neural networks (CNN) and recurrent neural networks (RNN), the spectrum of machine learning tools available underscores their versatility in tackling the dynamic nature of phishing threats.

An essential aspect of recent research lies in the exploration of feature engineering mechanisms, such as TF-IDF and word embedding, coupled with a diverse array of classifiers. These endeavors not only contribute to the nuanced understanding of email content and behaviors but also showcase the adaptability of machine learning models in discerning patterns indicative of phishing activities. The infusion of natural language processing (NLP) techniques further enriches the sophistication of anti-phishing strategies, enabling the models to comprehend the intricacies of human language and communication patterns.

The amalgamation of ensemble learning methods, including boosting and hybrid approaches, demonstrates a strategic synergy that enhances the robustness of anti-phishing solutions. The findings consistently indicate that ensemble machine learning methods, often combining the strengths of various classifiers, offer a significant leap forward in detection accuracy. This not only underscores the importance of diversity in model architecture but also emphasizes the need for ongoing innovation to stay ahead of increasingly sophisticated phishing tactics.

As we navigate the ever-evolving landscape of cyber threats, the holistic integration of machine learning into anti-phishing measures becomes imperative. Nevertheless, the journey does not end here. Continuous refinement, adaptation, and vigilance are crucial to ensuring the resilience of these systems. By fostering collaboration between researchers, industry professionals, and cybersecurity experts, we can collectively harness the potential of machine learning to fortify our defenses against phishing attacks and safeguard the integrity of digital communication. In essence, the future of anti-phishing lies in the persistent pursuit of innovation, drawing upon the rich tapestry of machine learning advancements to counteract the evolving challenges of the digital age.

In conclusion, the reviewed literatures establish that the

application of machine learning for anti-phishing measures has exhibited promising advancements in enhancing email security. The diverse range of techniques, from traditional models to deep learning algorithms, showcases the adaptability of machine learning in addressing the evolving challenges posed by phishing threats. As research continues to explore innovative approaches, it becomes evident that the fusion of natural language processing, ensemble learning, and deep-learning methodologies can significantly bolster the effectiveness of anti-phishing measures. However, ongoing efforts are crucial for staying ahead of sophisticated phishing tactics. The continuous refinement and integration of machine learning models into anti-phishing strategies are essential for creating robust defenses against ever-evolving cyber threats.

References

[1] F. Carroll, J. A. Adejobi, and R. Montasari, "How Good Are We at Detecting a Phishing Attack? Investigating the Evolving Phishing Attack Email and Why It Continues to Successfully Deceive Society," *SN Comput. Sci.*, vol. 3, no. 2, pp. 1–10, 2022.

[2] Statistica, "STATISTICA 2022," 2023. [Online]. Available: https://www.statista.com/statistics/266155/number-of-phishing-domain-names-worldwide/

[3] M. N. Rahim and K. P. M. Basheer, "A survey on anti-phishing techniques: From conventional methods to machine learning," *Malaya J. Mat.*, vol. 9, no. 1, pp. 319–328, 2021.

[4] Y. G. Zeng, "Identifying email threats using predictive analysis," in *2017 International Conference on Cyber Security and Protection of Digital Services, Cyber Security 2017*, London, UK., 2017, pp. 1–2.

[5] D. Rathee and S. Mann, "Detection of E-Mail Phishing Attacks – using Machine Learning and Deep Learning," *Int. J. Comput. Appl.*, vol. 183, no. 47, pp. 1–7, 2022.

[6] G. Kim Soon, C. Kim On, N. Mohd Rusli, T. Soo Fun, R. Alfred, and T. Tse Guan, "Comparison of simple feedforward neural network, recurrent neural network and ensemble neural networks in phishing detection," *J. Phys. Conf. Ser.*, vol. 1502, no. 1, 2020.

[7] E. Marková, T. Bajtoš, P. Sokol, and T. Mézešová, "Classification of Malicious Emails," in *2019 IEEE 15th International Scientific Conference on Informatics, Poprad, Slovakia*, Poprad, Slovakia, 2019, pp. 000279–000284.

[8] A. Sundararajan, G. Gressel, and K. Achuthan, "Feature selection for phishing detection with machine learning," *Int. J. Eng. Adv. Technol.*, vol. 8, no. 6 Special Issue 3, pp. 1039–1045, 2019.

[9] R. Basnet, S. Mukkamala, and A. H. Sung, "Detection of phishing attacks: A machine learning approach," *Stud. Fuzziness Soft Comput.*, vol. 226, pp. 373–383, 2008.

[10] P. Boyle and L. A. Shepherd, "MailTrout: A Machine Learning Browser Extension for Detecting Phishing Emails," *34th Br. Hum. Comput. Interact. Conf. Interact. Conf. BCS HCI 2021*, pp. 104–115, 2021.

[11] O. Ibitoye, R. Abou-Khamis, M. el Shehaby, A. Matrawy, and M. O. Shafiq, "The Threat of Adversarial Attacks on Machine Learning in Network Security: A Survey," *ArXiv*, 2023.

[12] P. Bountakas, K. Koutroumpouchos, and C. Xenakis, "A Comparison of Natural Language Processing and Machine Learning Methods for Phishing Email Detection," *ACM Int. Conf. Proceeding Ser.*, pp. 1–24, 2021.

[13] F. Toolan and J. Carthy, "Phishing detection using classifier ensembles," in *2009 eCrime Researchers Summit, eCRIME '09. Tacoma, WA, USA*, 2009, pp. 1–9.

[14] G. Apruzzese, M. Colajanni, L. Ferretti, A. Guido, and M. Mirco, "On the Effectiveness of Machine Learning and Deep Learning Algorithms for Cyber Security," in *Arc2018 10th International Conference on Cyber Conflict. Tallinn, Estonia.*, T. Minárik, R. Jakschis, and L. Lindström, Eds., Tallinn, 2018, pp. 371–390.

[15] P. Vaitkevicius and V. Marcinkevicius, "Comparison of Classification Algorithms for Detection of Phishing Websites," *Inform.*, vol. 31, no. 1, pp. 143–160, 2020.

[16] A. Basit, M. Zafar, A. R. Javed, and Z. Jalil, "A Novel Ensemble Machine Learning Method to Detect Phishing Attack," in *2020 IEEE 23rd International Multitopic Conference (INMIC), Bahawalpur, Pakistan.*, 2020, pp. 1–5.

[17] A. J. Wyner, M. Olson, J. Bleich, and D. Mease, "Explaining the success of adaboost and random forests as interpolating classifiers," *J. Mach. Learn. Res.*, vol. 18, pp. 1–33, 2017.

[18] I. B. Mustapha, S. Hasan, S. O. Olatunji, S. M. Shamsuddin, and A. Kazeem, "Effective Email Spam Detection System using Extreme Gradient Boosting." 2020.

[19] D. Anandita, Y. Pratap, P. Priyanka, K. Divya, and T. Rajesh, "A Novel Ensemble Based Identification of Phishing E-Mails," in *ICMLC '17: Proceedings of the 9th International Conference on Machine Learning and Computing, Singapore.*, 2017, pp. 447–451.

[20] S. S. M. M. Rahman, F. B. Rafiq, T. R. Toma, S. S. Hossain, and K. B. B. Biplob, "Performance Assessment of Multiple Machine Learning Classifiers for Detecting the Phishing URLs," *Advances in Intelligent Systems and Computing*, vol. 1079. pp. 285–296, 2020.

[21] M. Das, S. Saraswathi, R. Panda, A. K. Mishra, and A. K. Tripathy, "Exquisite Analysis of Popular Machine Learning–Based Phishing Detection Techniques for Cyber Systems," *Journal of Applied Security Research*, vol. 16, no. 4. pp. 538–562, 2021.

[22] K. L. Chiew, C. L. Tan, K. S. Wong, K. S. C. Yong, and W. K. Tiong, "A new hybrid ensemble feature selection framework for machine learning-based phishing detection system," *Inf. Sci. (Ny).*, vol. 484, pp. 153–166, May 2019.

[23] L. Gallo, A. Maiello, A. Botta, and G. Ventre, "2 Years in the anti-phishing group of a large company," *Comput. Secur.*, vol. 105, p. 102259, Jun. 2021.

[24] N. A. Unnithan, N. B. Harikrishnan, S. Akarsh, R. Vinayakumar, and K. P. Soman, "Machine learning based phishing E-mail detection," in *R. Verma, A. Das (eds.): Proceedings of the 1st AntiPhishing Shared Pilot at 4th ACM International Workshop on Security and Privacy Analytics (IWSPA 2018), Tempe, Arizona, USA.*, 2018, pp. 64–68. [Online]. Available: http://ceur-ws.org

[25] N. B. Harikrishnan, R. Vinayakumar, and K. Soman, "A machine learning approach towards phishing email detection," in *R. Verma, A. Das (eds.): Proceedings of the 1st AntiPhishing Shared Pilot at 4th ACM International Workshop on Security and Privacy Analytics (IWSPA 2018), Tempe, Arizona, USA.*, 2018, pp. 60–69.

[26] M. S. Munir Prince, A. Hasan, and F. Muhammad Shah, "A new ensemble model for phishing detection based on hybrid cumulative feature selection," *ISCAIE 2021 - IEEE 11th Symp. Comput. Appl. Ind. Electron.*, no. September, pp. 7–12, 2021.

[27] V. S. Mohan, J. R. Naveen, R. Vinayakumar, and K. P. Soman, "A.R.E.S: Automatic rogue email spotter," in *R. Verma, A. Das (eds.): Proceedings of the 1st AntiPhishing Shared Pilot at 4th ACM International Workshop on Security and Privacy Analytics (IWSPA 2018), Tempe, Arizona, USA.*, 2018, pp. 57–63. [Online]. Available: http://ceur-ws.org

[28] T. Gangavarapu, C. D. Jaidhar, and B. Chanduka, "Applicability of machine learning in spam and phishing email filtering: review and approaches," *Artif. Intell. Rev.*, vol. 53, no. 7, pp. 5019–5081, 2020.

[29] R. Vinayakumar, G. H. Barathi, K. M. Anand, K. Soman, and P. Prabaharan, "DeepAnti-PhishNet: Applying deep neural networks for phishing email detection," in *R. Verma, A. Das (eds.): Proceedings of the 1st AntiPhishing Shared Pilot at 4th ACM International Workshop on Security and Privacy Analytics (IWSPA 2018), Tempe, Arizona, USA.*, 2018, pp. 1–11. [Online]. Available: http://ceur-ws.org

[30] D. M. Ablel-Rheem, "Hybrid Feature Selection and Ensemble Learning Method for Spam Email Classification," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 1.4, pp. 217–223, 2020.

[31] B. Kitchenham and S. M. Charters, "Guidelines for performing Systematic Literature Reviews in Software Engineering," *ResearchGate*, no. October, 2021.

[32] A. El-Aassal, L. Moraes, S. Baki, A. Das, and R. Verma, "Evaluating Performance with New Metrics for Unbalanced Datasets," in *R. Verma, A. Das (eds.): Proceedings of the 1st AntiPhishing Shared Pilot at 4th ACM International Workshop on Security and Privacy Analytics (IWSPA 2018), Tempe, Arizona, USA.*, 2018, pp. 21–24. [Online]. Available: http://ceur-ws.org

[33] N. A. Unnithan, N. B. Harikrishnan, R. Vinayakumar, K. P. Soman, and S. Sundarakrishna, "Detecting phishing E-mail using machine learning techniques," in *Proceedings of the 1st AntiPhishing Shared Pilot at 4th ACM International Workshop on Security and Privacy Analytics (IWSPA 2018), Tempe, Arizona, USA.*, 2018, pp. 50–56. [Online]. Available: http://ceur-ws.org

[34] A. Vazhayil, N. B. Harikrishnan, R. Vinayakumar, and K. P. Soman, "PED-ML: Phishing email detection using classical machine learning techniques," in *R. Verma, A. Das (eds.): Proceedings of the 1st AntiPhishing Shared Pilot at 4th ACM International Workshop on Security and Privacy Analytics (IWSPA 2018), Tempe, Arizona, USA.*, 2018, pp. 69–76. [Online]. Available: http://ceur-ws.org

[35] F. Salahdine, Z. El Mrabet, and N. Kaabouch, "Phishing Attacks Detection A Machine Learning-Based Approach," in *2021 IEEE 12th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference, UEMCON 2023. New York USA.*, 2021, pp. 0250–0255.

[36] M. Hiransha, N. A. Unnithan, R. Vinayakumar, and K. P. Soman, "Deep learning based phishing E-mail detection," in *R. Verma, A. Das (eds.): Proceedings of the 1st AntiPhishing Shared Pilot at 4th ACM International Workshop on Security and Privacy Analytics (IWSPA 2018), Tempe, Arizona, USA.*, 2018, pp. 16–20. [Online]. Available: http://ceur-ws.org