Smishing Detection: Combating SMS Phishing Attacks by Utilizing Machine-Learning Algorithms

Aqsa Shaikh, Mariya Shaikh, Srivaramangai R.

***Abstract:** With the rapid uptake of mobile communications, cybercriminals have increasingly resorted to using SMS (Short Message Services) in the guise of phishing attacks commonly referred to as smishing (SMS phishing). Phishing SMS messages impersonate trusted organizations to persuade users into clicking malicious links, providing personal credentials, or installing malware. This paper reviews the latest advancements in machine learning for smishing detection, drawing on insights from various studies on the subject. It examines critical machine learning models, including Deep Learning models (CNN, LSTM), Logistic Regression, Random Forest, Support Vector Machines (SVM), and Gradient Boosting, to classify messages as spam, phishing, or legitimate. It examines feature extraction techniques such as TF-IDF, N-grams, and natural language processing (NLP) in the hope of improving detection accuracy. In this way, it also examines how cyber threat intelligence and real-world datasets, such as SpamAssassin, the UCI Machine Learning Repository, and PhishTank, can be utilised to develop robust models. The results show that ensemble learning and hybrid deep learning techniques are more effective at identifying objects than traditional methods, and they achieve this without increasing the number of false positives. Challenges such as adversarial SMS attacks, multilingual phishing messages, and limitations in real-time detection remain plausible. Future work needs to explore adaptability to real-time models, CTI-based threat analysis, and transparent AI (XAI) detection. Applying machine learning-driven smishing detection enhances the overall solution's intelligent, automated approach and adaptive defence mechanisms against evolving mobile phone phishing threats, resulting in increased security for mobile devices and, consequently, their users.*

***Keywords:** Cyber Threat Intelligence, Machine Learning, Smishing Detection, SMS Spam Classification, Convolutional Neural Network, Long Short Term Memory, Term Frequency Inverse Document Frequency, Optimal Feature Extraction Algorithm, Independent Recurrent Neural Network, Capsule Network, Random Forest*

Abbreviations:
SMS: Short Message Service
SVM: Support Vector Machines
CTI: Cyber Threat Intelligence
ML: Machine Learning
LSTM: Long Short-Term Memory
TD-IDF: Term Frequency Inverse Document Frequency
OFVA: Optimal Feature Extraction Algorithm
IndRNN: Independent Recurrent Neural Network
CapsNet: Capsule Network
RF: Random Forest
CTI-MURLD: Cyber Threat Intelligence-based Malicious URL Detection
LSTM: Long Short-Term Memory
BoW: Bag-of-Words
MLP: Multilayer Perceptron
CNN: Convolutional Neural Networks
AI: Artificial Intelligence
URL: Uniform Resource Locator
MTD: Mobile Threat Defence
QML: Qt Modelling Language
NLP: Natural Language Processing

Manuscript received on 15 February 2025 | First Revised Manuscript received on 18 February 2025 | Second Revised Manuscript received on 01 April 2025 | Manuscript Accepted on 15 April 2025 | Manuscript published on 30 April 2025.
*Correspondence Author(s)

Aqsa Shaikh, Student, Department of Information Technology, University of Mumbai, Mumbai (Maharashtra), India. Email ID: skaqsa242@gmail.com

Mariya Shaikh, Student, Department of Information Technology, University of Mumbai, Mumbai (Maharashtra), India. Email ID: shaikhmariya2909@gmail.com

Srivaramangai R.*, Head of the Department of Information Technology, University of Mumbai, Mumbai (Maharashtra), India. Email ID: rsrimangai@gmail.com, ORCID ID: 0000-0003-2723-6067

I. INTRODUCTION

The onslaught of scamming is inescapable, what with us using our phones for everything these days. Smishing (SMS phishing) is one of the scariest threats. This involves attackers sending fraudulent messages that appear to come from trusted organisations, such as banks, delivery services, and government agencies. Such short messages either harass the receivers into clicking on malicious links or solicit sensitive information that opens a door to financial fraud, identity theft, or even malware infection. There is a need to detect such scams as they become more sophisticated over time. Conventional smishing detection approaches, such as blacklisting and rule-based filters, are no longer practical since attackers have learned to update their arsenal continually. Machine learning (ML) algorithms are now being employed as robust tools to combat the significant issue of smishing message detection. ML models can learn from past evidence to identify patterns in smishing messages and detect suspicious ones, all while enhancing their capabilities through continuous learning.

Therefore, the main emphasis of this paper is on:

A. How machine learning can help in the detection of smishing messages.
B. A comparison of different ML models and their relative performance in detecting phishing.
C. The issue of detecting new attacks and reducing false positives.
D. Recommendations for the future development of AI-based SMS phishing detection with real-time data analysis.

The work is concerned with using "machine learning," a sub-parcelling of AI incorporation, in the detection of smishing messages. This paper provides a detailed review of various machine learning methods for smishing detection. The proposed methods include neural networks, support vector machines (SVMs), deep learning models, random forests, and decision trees for comparing their efficiencies in distinguishing phishing messages. The paper discusses various methods of feature extraction, including text-based methods (TF-IDF, N-grams), URL analysis, and behavioural analysis, which contribute to improving the accuracy of detection levels. We will discuss various machine-learning methods and their applications in combating such text message schemes. In doing so, we aim to give a detailed overview of how this technology can be used against smishing and what challenges remain.

II. LITERATURE REVIEW

According to Timko et al. [1], the study evaluated users' ability to differentiate between genuine messages and smishing messages through an online survey involving 187 participants. The implications of the findings suggest that attention and security behavioural scores significantly affect users' accuracy when identifying smishing messages. The author states that the accuracy of fake messages was 67.1%, while the accuracy of real messages was 43.6%, indicating the difficulty in the matter concerning user awareness. Mahmood and Hameed [2] note an increasing threat of smishing, which involves phishing via SMS messages in the mobile communication sector. The study proposes a method to detect smishing that utilises a combination of URL inspection by Google Engine and VirusTotal, along with an analysis of the content in the SMS. The research employed four machine-learning classifiers, building on the classification of messages into legitimate versus smishing, using Support Vector Machine (SVM), Random Forest (RF), Adaptive Boosting (AdaBoost), and Extreme Gradient Boosting (XGBoost). The model thereby attained an accuracy of 98.5%, which is superior to existing methods concerning smishing message detection. Ankit et al. [3] note that smishing—a growing threat to Cybersecurity—deceives users into installing malicious software. The approach is a two-phase machine learning methodology for spam and smishing text detection: first, spam is differentiated from ham, and subsequently, smishing detection is performed on spam. Achieving 96% accuracy with different classifiers, notably neural networks. The study confirms the importance of unique phishing features and information gain values for improved classification quality. Sharif et al. [4] note that detection of suspicious text is a vital subject in cybersecurity, particularly in Bengali language processing. In his studies, he proposed a machine-learning-based model for classifying text as either suspicious or non-suspicious. With a collection of 7,000 Bengali text documents, the model, which encompasses classifiers such as logistic regression and random forest, achieved an accuracy of 84.57%. His study emphasizes that both unigram and bigram features are significant for enhancing classification performance. Mohanty et al. [5] identify that identifying cyber threats posed by suspicious URLs is crucial for internet security. It also creates unique models of multi-class ML-based classification to recognise instances of spam, phishing, defacement attacks, and malware. A model utilising classifiers such as decision trees and logistic regression can achieve an accuracy of 94.1% by leveraging lexical features of URLs. It is demonstrated that feature extraction and classification techniques can significantly enhance the current state of threat detection. Sohn et al. [6] have taken spam filtering a step further by incorporating stylistic features beyond plain content-based filtering. This study examines the effectiveness of stylometric features, including the use of special characters, word length, part-of-speech n-grams, and frequency of function words, in distinguishing between spam and legitimate messages. Using a machine-learning model, their method enhances detection accuracy by minimising false negatives while maintaining a very low false-positive rate. Thus, the results indicate that the inclusion of writing style patterns has considerably improved the performance of SMS spam filters. Schlette et al. [7] note that Cyber Threat Intelligence (CTI) is crucial for managing security incident responses and their associated threat data in an organised manner within an organisation. The study evaluates the six CTI formats identified as key concepts for increased efficiency in incident response. This involves discussing playbooks, automation, and standardising the format in cybersecurity defence. The results suggest that using multiple CTI formats enhances response capability, enabling organisations to mitigate cyber threats more effectively. Chong et al. [8] note that malicious URL detection is necessary because the increasing use of mobile devices has led to the evolution of web vulnerabilities. Their efforts involve combining machine learning with URL lexical features, JavaScript source code features, and payload size to identify malicious URLs. Using an SVM with a polynomial kernel, they achieved 81% accuracy and a 74% F1 score. The work discusses the reassurance provided by real-time detection of malicious URLs, highlighting how features such as URL patterns and JavaScript obfuscation markers can significantly bolster security. Jain et al. [9] note that spam and phishing attacks are critical threats in cybersecurity. His work proposes machine learning-based detection of spam messages and phishing links. The model is built using datasets of known phishing sites and spam messages for newly incoming content, which are classified based on unique features. Various algorithms, including Random Forest and Support Vector Classifier, were evaluated and exhibited high capabilities in detecting threats. Jaiswal and Raut [10] note that cybersecurity presents URLs as a significant threat due to scams, data theft, and the spread of malicious programs. Hence, the study suggests machine learning to detect and block URLs. The proposed model enhances detection by analysing content through web crawling and sentiment analysis. The study highlights a pressing need for enhanced detection methodologies, as the changes in the cyber world are also evolving the threats. Aljabri et al. [11] have noted that malicious URLs pose an increasing threat to cybersecurity, resulting in phishing, malware, and data breaches. The review examines machine learning techniques for detecting these threats, analysing various types of features, detection methods, and the limitations of the datasets. The study argues that traditional blacklists are unable to identify new threats, and that supervised learning and deep learning models show promise in detecting malicious URLs. The paper also discusses the challenges of dataset quality and the evolving nature of attacks,

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)
© Copyright: All rights reserved.

emphasising the need for detection techniques that can learn and adapt to these changes. Tamal et al. [12] note that the growing nature of phishing attacks poses an ever-increasing cybersecurity threat due to their evolving nature and the inertia of existing detection techniques. This study aims to present the OFVA (Optimal Feature Extraction Algorithm) as a potential means of improving the existing phishing detection using supervised machine learning (ML). After analysing 274,446 URLs, 41 important intra-URL features were extracted from the model, and multiple classifiers were applied; among these, random forests achieved an accuracy of 97.52%. The study argues for the need for adaptive detection methods to wage an efficient war against phishing attacks. Rifah et al. [13] note that URL detection is crucial in safeguarding against a spectrum of cyber threats, including phishing and malware attacks. They propose a machine learning detection framework based on logistic regression for classifying URLs as safe or unsafe. The system relies on analysing URL structure and behaviour to effectively identify links that may pose a threat, without relying on webpage content. This method protects cyberspace via rapid and effective threat detection. Li and Dib [14], propose a new machine-learning architecture for detecting both known and unknown malicious URLs in real-time. The system classifies URLs into three major classes: phishing, malware, and others, utilising tree-based algorithms and CL-K-means. It achieves 92.54% accuracy in detecting zero-day attacks, completing a single classification in under 14 milliseconds. Yuan et al. [15] proposed a model for parallel neural joint detection of malicious URLs. The above system converts URLs into word embeddings and grayscale images, then extracts semantic and visual features and processes them with IndRNN and CapsNet with an attention mechanism. This achieves a very high classification accuracy and outperforms traditional techniques. Xuan et al. [16] propose a machine learning-based method to detect malignant URLs. The system extracts features from lexical features, host-based features, and correlated groups to create new URL attributes for training Random Forest, SVM, etc., to categorise them as safe or malignant. The outcome demonstrated that the Random Forest Model can achieve high accuracy on larger datasets, making this approach not only efficient but also practical. Ravindra et al. [17] note that phishing poses a significant security threat in cybersecurity by deceiving users into divulging sensitive information. The study presents a Random Forest machine learning-based detection system that classifies a URL as either legitimate or phishing. The system analyses URL features, including length, suspicious characters, and subdomains, to enhance detection accuracy. With a dataset of 4000 URLs, this model has achieved an accuracy of 86%, providing effective identification and blocking of phishing threats. Cao and Caverlee [18] Note That Social media spam URL detection is important in mitigating phishing and malware attacks. The research describes a type of detection based on behaviour that examines the posting patterns of URLs, along with user interactions with those URLs. By measuring and evaluating fifteen behavioural features using a dataset of seven million URLs, the approach achieves the highest accuracy, with 86% precision and recall. This method strengthens spam detection beyond the traditional techniques that rely on blacklists. Reyes-Dorta et al. [19] evaluated ML and QML techniques for detecting fraudulent URLs in their study. More than 90% true positive rates were achieved using classical ML algorithms like decision trees, logistic regression, and neural networks. Subsequently, QML was explored through methods such as the Variational Quantum Classifier, yielding promising results that seem to match those achieved by classical models. The authors highlight the promise of QML in cybersecurity, albeit limited due to the constraints of current quantum hardware and datasets. Sonowal [20] suggests that improving smishing message (SMS phishing detection) can be achieved through feature selection and machine learning. Conducted five different ranking algorithms and concluded that the one with the best performance is Kendall rank correlation with an AdaBoost classifier, which showed the highest accuracy of 98.40% by reducing the feature set to 61.53%. Thus, indicating an efficient method for improving smishing detection. Canali et al. [21]. In essence, the goal of Prophiler is to quickly filter out malicious web pages in the most efficient manner possible. This system utilises static analysis techniques to identify malicious features within the HTML content, JavaScript code, and URL structure of a specific site. The classification mechanism distinguishes between benign and malicious web pages using machine learning classifiers. Prophiler significantly reduces the workload of dynamic analysis tools by filtering approximately 85% of benign pages. The study demonstrates how Prophiler achieves high accuracy in detecting actual threats while maintaining a low false negative rate. This makes it truly usable for large-scale web security. Tabassum et al. [22] note that malicious URLs pose threats to cybersecurity due to phishing websites and malware. Traditional blacklisting methods have proven to be ineffectively slow at best when it comes to addressing new threats. It is against this backdrop that the study investigates machine learning for better accuracy in the detection of harmful URLs. Its findings indicate that Random Forest and Neural Networks achieve an accuracy of above 90% in various tests. Regarding the research's findings, it also highlighted the evolving nature of threats and challenges posed by zero-day attacks, as well as proposed adaptations for future improvements. Ghaleb et al. [23] Note That Malicious websites pose a significant cybersecurity threat, often used for phishing, malware, and fraud. It's extremely challenging to detect new types of attacks using traditional detection approaches. In this paper, we present Cyber Threat Intelligence-based Malicious URL Detection (CTI-MURLD), an ensemble learning-based model to improve detection accuracy. It extracts URL-based cyber threat intelligence (CTI) and Whois-based features based on a web search. The best classification improves detection accuracy through a two-stage classification process that combines Random Forest (RF) and Multilayer Perceptron (MLP). Improvement in accuracy is 7.8%, with a 6.7% decrease in false positives compared to previous methods. Das Gupta et al. [24] note that an increase in the instances of SMS being used as an avenue for fraud or sending unsolicited messages has made this study relevant. Conventional rule-based methods of detection often fail to deal with new spam patterns. This study proposes the use of machine learning to classify SMS as either spam or ham (legitimate). Several classifiers, including Random Forest (RF), Naïve Bayes, and Support Vector Machine (SVM), were evaluated to compare their accuracy in spam detection. The results showed that Random Forest achieved the best accuracy in classifying spam messages. The study points out that the success of spam detection is influenced significantly by feature selection and text preprocessing. Palwankar et al. [25], detected the malicious link to mitigate phishing and malware attacks. Traditional methods have become a hindrance, as they have been unsuccessful in keeping pace with emerging threats. This research proposes an approach based on machine learning, utilising logistic regression and random forest models. The system will be URL-based and not rely on the page's content. Improvements in operational efficiency will also involve integrating VirusTotal and GNews APIs for real-time verification and updates. The results yield improvements in both accuracy and security, making browsing safer.

III. OBSERVATIONS

A. Increasing Smishing Attacks Threatening Growth

Smishing/SMS Phishing—With serious cyber-attack potential, smishing also employs social engineering to coerce users into divulging sensitive information, downloading malware, or visiting malicious websites. In the modern world, where mobile communication is increasingly becoming an integral part of daily life, such cybercriminals rely more on trust in text messaging services rather than targeting individuals. Attackers often impersonate legitimate communications to appear as if they are from banks, governments, delivery services, or well-known brands, thereby establishing trust with users. Mobile banking and online transactions have only accelerated the ease with which financial scams, identity theft, or ransomware attacks can target new victims. Some mainstream rule-based detection methods, such as blacklists and keyword-based filtering, do not effectively capture the latest, evolving attack strategies. This fact renders them practically ineffective against so-called sophisticated phishing schemes.

B. Machine Learning Role in the Efficiency of Smishing Detection

Machine Learning (ML) is a powerful and streamlined mechanism for combating smishing cases, as it enables the use of automated and intelligent detection schemes that are adaptable to new threats. ML models, based on analysing message content, URLs, the sender's behaviour, and metadata, improve security through their in-depth scrutiny of suspicious messages. It is a reality that supervised learning models, such as support vector machines, random forests, decision trees, and neural networks, have consistently delivered highly accurate results in smishing detection. Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTMs) are deep learning models that have achieved excellent success rates in understanding language patterns and providing contextual information for messages, thereby positively impacting the classification of phishing messages. Combined with ensemble learning and deep learning, these hybrid approaches integrate different techniques, exhibiting superiority over unimodal approaches by reducing false positives and increasing detection rates. AI-backed, real-time dynamic detection systems, combined with reinforcement learning, are set to ensure dynamic protection by enriching knowledge bases with updated phishing patterns.

C. Feature Engineering and Data Analysis Significance

Extracting and analyzing relevant features extracted from text messages, URLs, and sender information is a fundamental aspect of improving detection rates for smishing.

- *Lexical Feature Extraction:* TF-IDF, N-grams, or Bag-of-Words (BoW) would perform extremely effective feature extraction, helping ML models analyse the structure of an SMS to detect abnormalities associated with phishing messages.
- *URL Inspection:* Analysis reveals details from IP address, domain age, and WHOIS information, utilising URL-based features that can also help classify fraudulent and abusive links embedded in SMS messages.
- *Behavioural Analysis: Tracking message patterns, response frequencies from senders, and behaviour will enable ML models to distinguish between potential phishing attempts and* simple communication.
- *Threat intelligence:* CTI—Integrating with external databases on cybersecurity threats from VirusTotal, PhishTank, and Google Safe Browsing will further enhance the ability of detection systems to identify malicious URLs or phishing attempts.

D. Challenges in Smishing Detection

Nonetheless, some challenges remain, even with the successful deployment of machine learning-based smishing detection.

- *Adversarial SMS Attacks:* Attackers constantly metamorphose their SMS messages to evade detection by ML. They employ various obfuscation techniques, including character replacement, addition of spaces, and word variation.
- *Multilingual Contextual Issues:* Many phishing messages are written in multiple languages, making it challenging for models to generalise well if they are trained on specific linguistic features.
- *Real-Time Processing Constraints:* Detect smishing messages in real-time while achieving high detection accuracy with low rates of false positives for security systems on mobile devices.
- *Privacy and Data Security Issues: For training machine learning models using SMS data, one needs to access user messages, which raises concerns about privacy and compliance with various data protection regulations, such as the GDPR.*

E. Future Research and Emerging Trends

Future research will focus on further developing security measures that are increasingly smarter, adaptable, and dynamic.

- *Explainable AI:* Making Interpretable decision-making in an ML model will help flag a message as malicious, allowing security teams and users to understand why something is flagged as malevolent.
- *Federated Learning for Detection and Preservation of Privacy: SMS data will not be stored in a centralised database for training; models will train locally on the user's device, making it safer than a centralised* system.
- *Adaptive and Self-Learning Models:* Reinforcement learning techniques will support continuous learning and adaptation to new phishing attack patterns.
- *Mobile Threat Defence (MTD): Smishing detection can then be combined with broader cyber strategies, including endpoint protection, fraud detection, and behavioural biometrics, to provide comprehensive* round-the-clock security solutions.
- *Automated Incident Response:* Integrating mobile security applications will speed up and automate the remediation of smishing.

IV. CONCLUSION

Smishing attacks pose a serious risk to cybersecurity, necessitating countermeasures that are far more advanced than traditional filtering techniques. Machine-learning-based approaches have demonstrated high accuracy in detecting smishing using deep learning, behavioural analysis, and real-time threat intelligence. Machine learning models, namely supervised, unsupervised, and deep learning-based techniques, have successfully found application in the detection and mitigation of smishing attacks. Several methodologies for feature extraction, including NLP, URL analysis, and behavioural monitoring, have improved specificity for false positives. Nevertheless, the issues posed by adversarial evasion attacks, multilingualism, and the need for real-time processing availability constrain continuous model improvement and research over time. The combination of explainable AI, cyber threat intelligence, and adaptive machine learning models constitutes the cornerstone for designing a robust, scalable, and transparent detection mechanism for smishing. By enhancing the research agenda on innovative threat detection, the cybersecurity fraternity can genuinely move the needle in mitigating the risk that SMS phishing poses and in improving mobile security overall.

DECLARATION STATEMENT

After aggregating input from all authors, I must verify the accuracy of the following information as the article's author.

- **Conflicts of Interest/Competing Interests:** Based on my understanding, this article does not have any conflicts of interest.
- **Funding Support:** This article has not been sponsored or funded by any organization or agency. The independence of this research is a crucial factor in affirming its impartiality, as it was conducted without any external influence.
- **Ethical Approval and Consent to Participate:** The data provided in this article is exempt from the requirement for ethical approval or participant consent.
- **Data Access Statement and Material Availability:** The adequate resources of this article are publicly accessible.
- **Author's Contributions:** The authorship of this article is contributed equally to all participating individuals.

REFERENCES

1. Daniel Timko, Daniel Hernandez Castillo, Muhammad Lutfor Rahman, "A Quantitative Study of SMS Phishing Detection," Unpublished Manuscript (arXiv Preprint), 2024, 16 pages, DOI: https://doi.org/10.48550/arXiv.2311.06911
2. Ameen R. Mahmood, Sarab M. Hameed, "A Smishing Detection Method Based on SMS Contents Analysis and URL Inspection Using Google Engine and VirusTotal," Iraqi Journal of Science, vol. 64, no. 10, 2023, 16 pages, DOI: http://doi.org/10.24996/ijs.2023.64.10.41
3. Ankit Kumar Jain, Sumit Kumar Yadav, Neelam Choudhary, "A Novel Approach to Detect Spam and Smishing SMS using Machine Learning Techniques," International Journal of E-Services and Mobile Applications, vol. 12, no. 1, January-March 2020, 21 pages, DOI: https://doi.org/10.4018/IJESMA.2020010102
4. Sharif Omar, Mohammed Moshiul Hoque, A. S. M. Kayes, Raza Nowrozy, and Iqbal H. Sarker, "Detecting Suspicious Texts using Machine Learning Techniques," Applied Sciences, vol. 10, no. 18, 2022, 23 pages, DOI: https://doi.org/10.3390/app10186527
5. Sanjukta Mohanty, Sourav Nanda, Rupayan Rout, Arpan Kumar, Vansam Agrawal, Arup Abhinna Acharya, Namita Panda, "Detection of Cyber Threats from Suspicious URLs Using Multi-Classification Approach" ResearchGate / Book Chapter, 2024, 14 pages, DOI: http://doi.org/10.4018/979-8-3693-1186-8.ch007
6. Dae-Neung Sohn, Jung-Tae Lee, and Hae-Chang Rim, "The Contribution of Stylistic Information to Content-Based Mobile Spam Filtering," Proceedings of the ACL-IJCNLP 2009 Conference Short Papers, 2009, 4 pages, https://aclanthology.org/P09-2081/
7. Daniel Schlette, Marco Caselli, and Gunther Pernul, "A Comparative Study on Cyber Threat Intelligence: The Security Incident Response Perspective," IEEE Communications Surveys & Tutorials, 2021, DOI: http://doi.org/10.1109/COMST.2021.3117338
8. Christophe Chong, Daniel Liu (Stanford), and Wonhong Lee (Neustar), "Malicious URL Detection", Unspecified publication, four pages, https://cs229.stanford.edu/proj2012/ChongLiu-MaliciousURLDetection.pdf
9. Ms. Shilpi Jain, Dr. Madhur Jain, Ridhi Kalia, Divyansh Rampal, "A Comprehensive Model for Spam Detection and Phishing Link Detection," International Journal of Scientific Research in Computer Science, Engineering and Information Technology, vol. 10, no. 3, 2024, 5 pages, DOI: http://doi.org/10.32628/CSEIT24103109
10. Muskaan V. Jaiswal and Anjali B. Raut, "Detecting and Blocking of Malicious URL," International Journal of Science and Research (IJSR), vol. 10, no. 6, 2021, 3 pages, DOI: http://doi.org/10.21275/SR21610230148
11. Malak Aljabri; Hanan S. Altamimi, Shahd A. Albelali, Maimunah Al-Harbi, Haya T. Alhuraib, Najd K. Alotaibi, "Detecting Malicious URLs Detection Using Machine Learning Techniques: Review and Research Directions," IEEE Access, vol. 10, 2022, 23 pages, DOI: http://doi.org/10.1109/ACCESS.2022.3222307
12. Maruf A. Tamal, Md K. Islam, Touhid Bhuiyan, Abdus Sattar, Nayem Uddin Prince, "Unveiling Suspicious Phishing Attacks: Enhancing Detection with an Optimal Feature Vectorization Algorithm and Supervised Machine Learning," Frontiers in Computer Science, vol. 6, no. 1428013, 2024, 16 pages, DOI: http://doi.org/10.3389/fcomp.2024.1428013
13. Amar Palwankar, Rifah Solkar, Afiya Borkar, Shreya Khedaskar, and Pranali Shingare, "Malicious Link Detection System," International Research Journal Engineering and Technology (IRJET), vol. 9, no. 11, 2022, 5 pages, https://www.irjet.net/archives/V9/i11/IRJET-V9I1165.pdf
14. Shiyun Li and Omar Dib, "Enhancing Online Security: A Novel Machine Learning Framework for Robust Detection of Known and Unknown Malicious URLs," Journal of Theoretical and Applied Electronic Commerce Research, vol. 19, no. 4, 2024, 42 pages, DOI: https://doi.org/10.3390/jtaer19040141
15. Yuan Jianting, Chen Guanxin, Tian Shengwei, Pei Xinjun, "Malicious URL Detection Based on a Parallel Neural Joint

Model," IEEE Access, vol. 9, 2021, 9 pages, DOI: http://doi.org/10.1109/ACCESS.2021.3049625

16. Cho Do Xuan, Hoa Dinh Nguyen, Tisenko Victor Nikolaevich, "Malicious URL Detection Based on Machine Learning," International Journal of Advanced Computer Science and Applications (IJACSA), vol. 11, no. 1, 2020, 6 pages, DOI: http://dx.doi.org/10.14569/IJACSA.2020.0110119
17. Salvi Siddhi Ravindra, Shah Juhi Sanjay, Shaikh Nausheenbanu Ahmed Gulzar, Khodke Pallavi, " Phishing Website Detection Based on URL," IJSRCSEIT, vol. 7, no. 3, 2021, 6 pages, DOI: https://doi.org/10.32628/CSEIT2173124
18. Cheng Cao, James Caverlee, "Detecting Spam URLs in Social Media via Behavioural Analysis," Lecture Notes in Computer Science (LNCS), Springer, vol. 9022, 2015, 12 pages, DOI: http://doi.org/10.1007/978-3-319-16354-3_77
19. Nuria Reyes-Dorta, Pino Caballero-Gil, Carlos Rosa-Remedios, "Detection of Malicious URLs Using Machine Learning," Wireless Networks, vol. 30, 2024, 18 pages, DOI: https://doi.org/10.1007/s11276-024-03700-w
20. Gunikhan Sonowal, "Detecting Phishing SMS Based on Multiple Correlation Algorithms," SN Computer Science, vol. 1, no. 361, 2020, 9 pages, DOI: https://doi.org/10.1007/s42979-020-00377-8
21. Davide Canali, Marco Cova, Giovanni Vigna, Christopher Kruegel, "Prophiler: A Fast Filter for the Large-Scale Detection of Malicious Web Pages," Proceedings of the 20th International Conference on World Wide Web (WWW 2011), 2011, 10 pages, DOI: https://doi.org/10.1145/1963405.1963436
22. Tasfia Tabassum, Md. Mahbubul Alam, Md. Sabbir Ejaz, Mohammad Kamrul Hasan, “A Review on Malicious URLs Detection Using Machine Learning Methods," Journal of Engineering Research and Reports, vol. 25, no. 12, 2023, 13 pages, DOI: http://doi.org/10.9734/JERR/2023/v25i121042
23. Fuad A. Ghaleb, Mohammed Alsaedi, Faisal Saeed, Jawad Ahmad, Mohammed Alasli, "Cyber Threat Intelligence-Based Malicious URL Detection Model Using Ensemble Learning," Sensors, vol. 22, no. 9, 2022, 19 pages, DOI: https://doi.org/10.3390/s22093373
24. Suparna Das Gupta et al., "SMS Spam Detection Using Machine Learning," Journal of Physics: Conference Series, vol. 1797, no. 1, 2021, 6 pages, DOI: http://doi.org/10.1088/1742-6596/1797/1/012017
25. Prof. Amar Palwankar, Afiya Borkar, Pranali Shingare, Rifah Solkar, Shreya Khedaskar, “Suspicious Link Detection Using AI," International Journal of Advanced Research in Science, Communication and Technology (IJARSCT), vol. 3, no. 3, 2023, 8 pages, DOI: http://doi.org/10.48175/IJARSCT-9171

AUTHOR’S PROFILE

Aqsa Shaikh is a student in the M.S. (Cyber Security) program at the University Department of Information Technology, University of Mumbai. Aqsa has been proactively involved in solving Hackathon problems. She is proficient in programming languages such as Java, Python, and jQuery, and has developed portals utilising MySQL as the backend. Her flair for Cybersecurity, with initial skills in tools like Wireshark and NMAP, led her to pursue a master's degree in cybersecurity, focusing on various types of attacks and detection. Her expertise includes Cryptography and Network Security, as well as Information Security. Currently, she is focusing on “Smishing Detection” as her research project area.

Mariya Shaikh is a student in the PG program, M.S. (Cyber Security), from the University Department of Information Technology, University of Mumbai. Mariya has good exposure to current trends in the IT industry, particularly the latest developments in cybersecurity. She has a varied skill set, ranging from MS SQL Server and PHP for web development to JSP and Python for research. She has actively participated in Hackathons and Project Exhibitions conducted by various educational institutions. She is currently focused on her research project for Smishing Detection and plans to pursue Cyber Security certifications. Her expertise includes Cryptography and Network Security, as well as Information Security.

Srivaramangai R., Head, Department of IT, University of Mumbai, India. Having 24 years of experience in teaching and 6 years in industry. The specialization areas are artificial intelligence, security, image processing. Has industry experience in web development and report code generators. Has published more than 35 International journal papers, 25 conference papers, and served as a resource person for various workshops, and chaired sessions. The papers relevant to Cyber Security includes “Assessment of Deep Packet Inspection System of Network traffic and Anomaly Detection”, Enhancing Security using ECC in Cloud Storage”, “Recapitulation of the Use of Machine Learning for Prevention of DDoS Attack on SDN Controller” and “Unmasking Deceptive Websites: Harnessing Machine Learning For Phishing Detection”.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of the Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)/ journal and/ or the editor(s). The Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP) and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.