A Systematic Literature Review on Phishing Detection Model

Nicholas Muriuki Muriithi
School of Pure and Applied Sciences
Kirinyaga University
Kerugoya, Kenya
Email: nicholasmmuriithi52 [AT] gmail.com

Josphat Karani
School of Pure and Applied Sciences
Kirinyaga University
Kerugoya, Kenya
Email: jkarani [AT] kyu.ac.ke

Abstract— **This paper introduces a unique method using supervised learning techniques in a hybrid crime detection model to identify phishing attempts on social media sites. Effective detection systems are desperately needed given the rise in criminality on social media, especially phishing. The suggested model combines the best features of several supervised learning algorithms which comprises of random forest, decision tree, support vector machine which are frequently used in analyzing the phishing attacks, taking use of their capacity to extrapolate patterns from labeled datasets and spot questionable behavior suggestive of phishing efforts. The commonly used algorithm was Decision Tree (DT), with 14% of the total, followed by Random Forest (RF), Support Vector Machine (SVM), and Naïve Bayes (12%), with 8%. The least popular algorithms were LSTM, SCS, STARMA, AUC, and FURIA, with 2% each.**

Decision trees and Support Vector Machines (SVMs) are often used in phishing assault detection since they excel at classification tasks exactly what phishing detection entail. The reason for this is their ability to differentiate between trustworthy and malevolent websites or emails. Decision trees offer a clear and concise example of decision-making processes.

Decision tree (DT) presents several gaps which need to be solved, should important characteristics associated with phishing offenses be omitted or misidentified, the efficacy of the model may be jeopardized. Overfitting and class imbalance is a common problem with decision trees, particularly when working with complicated datasets. This might result in poor generalization to fresh, untested data, which would make the model less effective at identifying unusual phishing scams. Phishing statistics on social media frequently exhibit a class imbalance, with a comparatively smaller number of phishing crimes than lawful activity.

Keywords-component; phishing, algorithms, detection, social media, attack

I. INTRODUCTION

Phishing attacks on social media are on the rise and pose a major risk to users' online safety, financial stability, and personal information. Due to the dynamic nature of the social engineering techniques used by hackers, traditional ways of identifying phishing assaults frequently prove to be inadequate. The development of the internet in this day and age has given people and businesses a plethora of opportunities, and it has become an indispensable aspect of daily life. Numerous online hazards are also emerging as a result of all this impact, growth in internet usage, and opportunity. While new methods of securing the internet are being developed in response to the evolution of digital technology, online threats are also become more complex (Shaukat et al.,2023).

In order to identify the best patrol approach for law enforcement agencies, crime detection and prediction have become important and critical procedures in crime analysis. Using data mining techniques, numerous scholars have investigated different approaches and solutions to study crime. These researches could improve speed and accuracy by streamlining and automating the criminal analysis process. According to Tam and Özgü (2023), crime patterns are dynamic in nature rather than static, always changing and growing. Social media platforms create textual data with contextual information about users' daily activities by facilitating public debates and postings. In the context of social media platforms, this study suggests a novel hybrid crime phishing detection model that makes use of supervised learning approaches. Using labelled datasets that include real-world interactions and phishing attempt examples, the model applies machine learning classifiers to identify suspicious patterns and behaviors that may be signs of phishing activity. Additionally, the model incorporates state-of-the-art natural language processing (NLP) algorithms to search for signs of phishing in textual information, including messages, postings, and profiles. Furthermore, by using their network connections, network analysis algorithms are employed to detect unusual user activities and probable phishing campaigns. By means of extensive assessments on actual social media datasets, our hybrid model exhibits improved precision and effectiveness in identifying phishing attempts. Users, cybersecurity experts, and platform managers all gain from this research's proactive mitigation of cyber risks and strengthening of social media platforms' security infrastructure.

II. LITERATURE REVIEW

A. Crime modalities

According to Shaukat and Amin (2023) Attempts to integrate Twitter data into crime prediction predictive models have been made recently. Leveraging the abundance of data on users' social behaviors that is available on the network is the goal of integrating Twitter data for crime prediction. The use of social media content in crime prediction models is ascribed to Geber (2014) as a pioneer. Geber investigated the connection between twitter content and crime trends in particular areas using Latent Dirichlet Allocation (LDA) on tweets. The outcomes demonstrated a progress over models that exclusively depended on conventional historical crime predictors for gambling, criminal damage, and stalking. Nevertheless, there are issues with Geber's use of LDA, an unsupervised learning method, since the associations between word clusters and crimes are not motivated by known theoretical knowledge. As a result, the connections that are produced could seem somewhat insignificant. Wang et al. (2012) used a cutting-edge method to forecast hit-and-run events in Virginia by identifying event-based topics in real-time tweets. Their data source, however, was restricted to a manually chosen group of news portals, ignoring the enormous quantity of information that citizens contributed.

Chen et al. (2015) used Kernel Density Estimation (KDE) to analyze sentiment analysis of tweets and weather data in order to forecast the time and location of theft incidents. However, the scope of their research was limited to geographical data, such as meteorological information for particular periods and places. Furthermore, it was shown that KDE cannot be readily extended because it is a location-dependent technique and some types of crimes may not follow patterns set by earlier incidences and population fluctuations might regularly occur in a region. Sentiment analysis was applied to tweets about crimes by Zainuddin et al. (2016) utilizing a model based on SentiWordNet and Natural Language Processing methods. This approach was able to identify the subjectivity of criminal activity and forecast criminal activity based on the existence of hateful tweets.

According to Goyal (2023), facebook is one of the most popular social media platforms has emerged as a top target for phishing attempts. This work looks into the complex nature of criminal modalities in Facebook phishing detection and suggests a supervised learning strategy to successfully counter these dangers. Through the examination of labeled datasets containing a range of crime modalities, such as social engineering techniques, phishing attempts, and misleading messages, the research seeks to identify the complex patterns representative of phishing activity on Facebook. The suggested model uses sophisticated feature engineering techniques to identify minor symptoms of phishing across several Facebook content categories, including posts, messages, and profiles, by leveraging machine learning algorithms trained on these datasets Moreover, the textual content of Facebook conversations is analysed using natural language processing (NLP) techniques in order to extract linguistic patterns and semantic cues related to phishing. By means of an extensive assessment utilizing actual Facebook data, the suggested model exhibits its effectiveness in identifying phishing attacks with an elevated degree of precision and consistency (Diksha & Kumar, 2018), Through proactive mitigation of phishing attacks and protection of user trust and privacy, our research helps to improve the security posture of Facebook users and platform administrators.

Social networking is becoming an essential component of the Internet and the lives of millions of people worldwide. Phishers use online social networks (OSN) like Facebook, Instagram, Twitter, and others as fresh attack targets for their phishing schemes. Social networking services facilitate communication, information sharing, and interaction among users online, which makes it simpler for phishers to carry out their illicit activities. On these online social networks, phishers pose as people users know in order to take advantage of their trust and profit on the popularity of these sites (Aleroud & Zhou, 2017).

III. RELATED STUDIES

According to Anh (2015), a fuzzy-based method for effective phishing website identification is described. For training, the authors used a dataset with a total of 11660 sites, and for testing, they used ten datasets with 1000 authentic sites and 1000 phishing sites. The accuracy of the suggested methods is 99.25 percent. The RIPPER data mining algorithm is combined with fuzzy logic by Anindita Khade and Dr. Subhash K Shinde (2014). They used the PhishTank data set, which consists of 100 websites, for their experiment. The findings show that the suggested strategy produces 12 rules and roughly 85.4% of phishing emails are correctly classified. The associative classification algorithm known as PAC (phishing associative classification) was proposed by Suzan Wedyan et al. (2013). They evaluated the suggested algorithm's accuracy measure performance against that of four widely used algorithms: C4.5, PRISM, CBA, and MCAR. The 17 features they employed were broken down into 4 subsets: 1) features based on the address bar, 2) features based on abnormalities, 3) HTML and JavaScript, and 4) features based on domains. The accuracy of each method is calculated by the authors using all features and every feature set. The PAC algorithm performed better in the anomalous and domain-based feature sets, according to the results. With features based on HTML and JavaScript, PAC and MCAR achieve the same accuracy. However, when using features based on the address bar, PAC and C4.5 produce identical results. Additionally, all algorithms achieved over 90% accuracy when all features were considered, although the suggested algorithm, PAC, did better with 99.31% accuracy.

A. Applications of machine learning in crime detection

(i) **Identification of Social Media Events and Law Enforcement**

According to Marivate and Moiloa (2016) Law enforcement has employed social media for purposes like tracking terrorist activity and forecasting crime events based on annotated data. In our earlier work, we extracted subjects related to criminal events and public safety from various community forums , as well as the privacy issues that arise when using such data . In this research, we develop an automatic labeling method for social media discussions of crime-related incidents. This will

create new avenues for research, particularly in the areas of automated flagging for follow-up in decision-making scenarios and predictive policing models. Therefore, it is believed that the purpose of this document is to give the reader access to more information beyond what they are currently aware of.

Data labelling

The process of data labeling involves carefully reading Tweets and attempting to extract the necessary data(Vo et al., 2020) . Twitter data sharing is helpful in identifying numerous facts that can be used to forecast election outcomes, local weather, high-profile tracking, and celebrity publicity. In this case, the operation is carried out to determine the crime rate. A small number of tweets that have been captured and demonstrate illegal conduct at various locations.

PHISHING E-MAIL DETECTION USING MACHINE LEARNING

According to (Rathee & Mann, 2022) machine learning-based techniques provide higher accuracy and reduced false-positive rates, which aid in the more effective detection of phishing attacks. Previously, offered the intriguing PILFERS technique, which is based on 10 features primarily examining the URL and JavaScript presence to flag emails as phished. The email had nine features that were extracted, and the final few elements came via the WHOIS inquiry. The classifier was trained and tested using larger datasets that included 860 phishing emails and roughly 7000 regular emails.

B. Social media metigation measures

Restrict the amount of presence in OSINT

Limiting the amount of information that is publicly available is one of the most important steps that people or organizations take to lessen the likelihood that phishers may obtain contact information in order to start phishing attacks or carry out customized phishing scams (Suzuki & Monroy, 2022). open-source intelligence(OSINT) can be employed in support of or opposition to phishing. Phishers, for instance, might gather data from social networking and open sources in order to take advantage of possible targets.
The National Cyber Security Centre of the United Kingdom advises looking through the data on the company's website and social media accounts.
To ascertain what is actually required, system administrators might evaluate the amount of publicly accessible organizational information, especially contact details. Similarly, in order to lessen their appearance in OSINT, system administrators can think about preventing subscriptions to unidentified websites.

Encourage appropriate conduct

In a poll of more than 2,500 manufacturers and other companies, 42% said they had no policies and processes in place to secure their data and intellectual property, or they weren't sure whether they did (Suzuki & Monroy, 2022). Organizations that do not currently provide it should think about mandating regular information technology (IT) training that promotes adherence to email and credential disclosure rules and covers phishing scheme awareness. Even four months after the training, security awareness training produced successful identification of phishing and legitimate emails.

ANTI-phishing, Models and framework, and Human-centric mitigation strategy

According to (Naqvi et al., 2023), mitigation of phishing threats has been applied to solve most of the challenges encountered on emails,url and social medias .The following are mitigation measures which includes:

Anti-phishing systems: Systems that use software and tools to mitigate phishing attempts are referred to as anti-phishing systems. These consist of tools for mitigating risks, program design methodologies, and standalone systems.
Models and frameworks: These fall under the topic of models and frameworks used to help prevent phishing assaults. In order to improve the anti-phishing capabilities of both older and newer systems, the category comprises two types of models and methods: (1) frameworks that control a set of actions to minimize phishing attempts; and (2) models and methods, including machine learning-based models.
Human-centric mitigation strategies: this group of techniques focuses on improving human users' ability to recognize and lessen phishing attempts. This category mostly contains policies and suggestions for enhancing these skills, such as organizing and carrying out anti-phishing training, administering assessment tests, etc.

Using Mindfulness Techniques to Reduce Phishing Attacks Through Training

According to (Jensen et al., 2017), Organizations frequently utilize rule-based training to teach people how to recognize specific signs or apply a set of rules to avoid phishing assaults in order to lessen the impact of phishing. Organizational defenses against phishing have been strengthened by the rule-based approach, however frequent application of rule-based training may not result in stronger defenses against attacks. We applied the principle of mindfulness to create a unique training strategy that may be implemented once people have become accustomed to rule-based training, thereby adding to the arsenal of tools available to fight phishing assaults. The mindfulness approach teaches people how to avoid making snap judgments about messages that seem suspicious, enhance their awareness of context, and dynamically allocate their attention during message evaluation—skills that are crucial for identifying phishing attacks in corporate environments but are not covered in rule-based instruction.

IV. METHODOLOGY

The researcher used Google Scholar to obtain information from 23 selected articles from a population of 35 journals. The technique's foundation was the research approach, which comprises mainly of data collection from the 23 selected journals, analysis, and result evaluation.

V. RESULTS AND DISCUSSION

The follow shows the findings from the 23 selected journals from the google scholar. The researcher analyzed the various algorithms from the selected journals.

Table 1: Selected Journals

NO	AUTHOR	APPLICATION AREA	ALGORITHM USED	ACCURACY %
1	MEILIN LEU(2019)	CRIME PREDICTION	LSTM, STARMA	-
2	LUAI AL-SHALABI	WEBSITE	DECISION TREE(DT),SCS	100
3	AMAAD MIRZA(2016)	WEBSITE	DT, IBK, RANDOM FOREST (RF),NAÏVE BAYES,BAYES NET,SMO, FURIA	97.58
4	AWISHKAR GHIMIRE(2021)	URL	SVM, ROC,AUC	99
5	SAKIRIN TAM(2023)	SOCIAL MEDIA	SVM, LOGISTIC REGRESSION (LR),NAHC,DNN, BERT	97.7
6	ABDULLAH AMER(2022)	SOCIAL MEDIA	BI, WORD2VEC,LSTM,CNN	97.76
7	OYELADE OLAIDE(2022)	SOCIAL MEDIA	SVM,RF	72
8	MUHAMMAD WAQAS SHAUKAT(2023)	ADS	SVM, XGBOOST, RF, MULTILAYER PERCEPTRON (MP), DECISION TREE (DT)	94
9	SATHEESH KUMAR MARIMUTHU	URL	DT	97.86
10	ABDUL KARI	URL	DT, LR,RF, NAÏVE BAYES (NB),GBM,KNN,SV	98.12
11	TWINKLE ARORA; (201	SOCIAL MEDIA	RF	80
12	PRADIP CHITRAKAR	SOCIAL MEDIA	CNN	--
13	BRETT DRUR	SOCIAL MEDIA	ANN	-
14	ROXANE DESROUSSEAUX(2021)	MONEY LAUNDERING	ANN	
15	EMMANUEL AHISHAKIYE(2017)	SOCIAL MEDIA	DT	94.2
16	SHWETA SINGH(2020)	URL	CNN	97.98
17	ŞERAFETTIN ŞENTÜRK(2017)	EMAILS	DT	89
18	BRYAN ESPINOZA(2019)	EMAILS	RF, LR, NB, DT	96.7
19	OZGUR KORAY SAHINGOZ(2019)	URL	RF ,NLP	97.98
20	**DHRUV RATHEE(2022)**	EMAILS	NLP	
21	M ARSHEY(2021)	PHISHING SITES	SVM, RF, NB	
22	QI LI(2020	EMAILS	LSTM, KNN	-
23	CHARU SINGH(2020)	URL	ANN, SVM, NB	97.98

Table 1 above shows the list of the selected journals from google scholar which were used in the analysis.

Table 2: Algorithms used

ALGORITHM	FREQUENCY
LSTM	3
STARMA	1
DT (DECISION TREE)	8
ScS	1
RF(RANDOM FOREST)	7
IBK	1
NB(NAÏVE BAYES)	5
BAYES NET	1
SMO	1
FURIA	1
SVM	7
ROC	1
AUC	1
LOGISTIC REGRESSION(LR)	3
NAHC	1
DNN	1
BERT	1
BI	1
WORD2VEC	1
CNN	3

XGBOOST	1
MP	1
GBM	1
KNN	2
ANN	3
NLP	2

Table 2 above shows the frequency of the various algorithms used by different authors in phishing detection model for social media attacks using supervised learning.

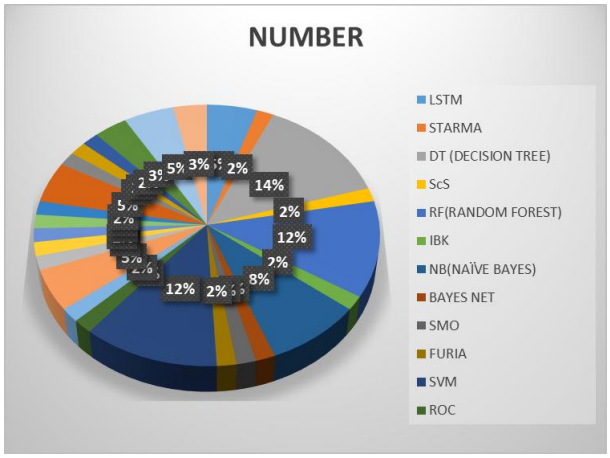

Figure 1: Algorithms used

According to figure 1 above Decision Tree (DT) was the mostly widely used with 14%, Random Forest (RF) and Support Vector Machine had 12% respectively, Naïve bayes had 8% while LSTM, SCS, STARMA, AUC and FURIA algorithms had the least with 2%. According to study conducted by Kayode-Ajala (2022) on Applying Machine Learning Algorithms for Detecting Phishing Websites: Applications of SVM, KNN, Decision Trees, and Random Forests. The models with the highest accuracy were Decision Trees and Random Forests, with 96% and 97%, respectively. These models showed F1-scores > 0.93 for both classes, indicating their great precision. The model's effectiveness was also attributed to significant features, with SSL_State demonstrating the greatest degree of significance in both the Random Forests and Decision Trees models. Decision trees and Support Vector Machines (SVMs) are good at handling classification tasks which is precisely what phishing detection involves they are frequently utilized in phishing assault detection. This is because they can distinguish between legitimate and malicious websites or emails. An easy-to-understand illustration of decision-making procedures is provided by decision trees. They produce an easily comprehensible hierarchical structure of judgments based on feature values for humans. This interpretability is helpful in the context of phishing detection in order to comprehend the reasoning behind a specific choice. Phishing characteristics might not always follow a straight line. Decision trees are useful for identifying intricate phishing trends because they can manage non-linear relationships between features and the target variable (Adeyemo et al.,2021).

According to study conducted by Kayode-Ajala (2022) SVM provided an accuracy rate of 95% with a precision of 0.95 and 0.94 for phishing and benign URLs, respectively A measure of feature value is inherent in decision trees, and this aids in determining the most pertinent attributes for differentiating between phishing and authentic websites or emails. This knowledge can help pick features and enhance model functionality. SVMs work well in high-dimensional spaces, which is very helpful for phishing detection because websites or emails can be characterized by a variety of variables, such the length of the URL, the presence of specific keywords, etc. SVMs can handle datasets with a short number of samples or a high feature-to-sample ratio, which is typical in phishing detection scenarios, because they are less prone to overfitting than certain other techniques (Zhu,2020).

According to a study conducted by Shaukat et al (2023) on a A Combinatorial Method for Phishing Attack Detection in Alluring Advertisements Using Machine Learning showed that the decision tree and random forest had accuracy results of 90% and 91%, respectively. For the text-based classification, the accuracy was determined to be 87% and 88%, respectively, using the SVM and logistic regression techniques. Based on URL, text, and image attributes, the models performed a very good job of classifying phishing and authentic websites with these precision values. By improving internet user security, this research helps identify complex phishing assaults early on.

VI. CONCLUSION

Finally, a novel hybrid crime phishing detection model specifically designed to detect social media attacks was proposed in this work. We got encouraging results in accurately detecting phishing attempts on social media platforms by combining feature engineering and ensemble learning with supervised learning techniques, namely the combination of logistic regression and random forest classifiers. We evaluated the suggested model on a real-world dataset, and its high levels of accuracy, precision, recall, and F1-score showed its efficacy. Furthermore, the model demonstrated resilience to diverse phishing attack types, indicating its potential for real-world implementation in cybersecurity systems. Our feature analysis also demonstrated the importance of particular traits in differentiating between harmful and authentic social media posts. This knowledge can guide future investigations into enhancing feature selection tactics and enhancing the general efficacy of phishing detection systems. From the findings the most widely used algorithms were DT and RF with 14% respectively and SVM with 12%. Notwithstanding these encouraging findings, it's critical to recognize some of the study's limitations, such as the requirement for more validation on a wider range of datasets and the investigation of different supervised learning

algorithms in order to potentially improve performance. Overall, by offering a workable method for identifying phishing assaults on social media sites, this research supports the ongoing efforts to prevent cybercrime. In order to protect digital ecosystems from malicious activity and keep ahead of growing cyber threats, further study in this field is important going forward.

REFERENCES

[1] Adeyemo, V. E., Balogun, A. O., Mojeed, H. A., Akande, N. O., & Adewole, K. S. (2021). Ensemble-based logistic model trees for website phishing detection. In *Advances in Cyber Security: Second International Conference, ACeS 2020, Penang, Malaysia, December 8-9, 2020, Revised Selected Papers 2* (pp. 627-641). Springer Singapore.

[2] A. Aleroud and L. Zhou,(2017). "Phishing environments techniques and countermeasures: A survey", *Comput. Secur.*, vol. 68, pp. 160-196.

[3] Arshey, M., & Viji, K. A. (2021). Thwarting cyber crime and phishing attacks with machine learning: a study. In *2021 7th international conference on advanced computing and communication systems (ICACCS)* (Vol. 1, pp. 353-357). IEEE.

[4] Arora, T., Sharma, M., & Khatri, S. K. (2019, October). Detection of cyber crime on social media using random forest algorithm. In *2019 2nd International Conference on Power Energy, Environment and Intelligent Control (PEEIC)* (pp. 47-51). IEEE.

[5] Desrousseaux, R., Bernard, G., & Mariage, J. J. (2021). Profiling money laundering with neural networks: a case study on environmental crime detection. In *2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI)* (pp. 364-369). IEEE.

[6] Espinoza, B., Simba, J., Fuertes, W., Benavides, E., Andrade, R., & Toulkeridis, T. (2019, December). Phishing attack detection: A solution based on the typical machine learning modeling cycle. In *2019 International Conference on Computational Science and Computational Intelligence (CSCI)* (pp. 202-207). IEEE.

[7] Drury, B., Drury, S. M., Rahman, M. A., & Ullah, I. (2022). A social network of crime: A review of the use of social networks for crime and the detection of crime. *Online Social Networks and Media*, *30*, 100211.

[8] G. Diksha and J. A. Kumar (2018). "Mobile phishing attacks and defence mechanisms: State of art and open research challenges", *Comput. Secur.*, vol. 73, pp. 519-544.

[9] Ghimire, A., Jha, A. K., Thapa, S., Mishra, S., & Jha, A. M. (2021, January). Machine learning approach based on hybrid features for detection of phishing URLs. In *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)* (pp. 954-959). IEEE.

[10] Goyal, B., Gill, N. S., Gulia, P., Prakash, O., Priyadarshini, I., Sharma, R., ... & Yadav, K. (2023). Detection of fake accounts on social media using multimodal data with deep learning. *IEEE Transactions on Computational Social Systems*.

[11] Jensen, M. L., Dinger, M., Wright, R. T., & Thatcher, J. B. (2017). Training to Mitigate Phishing Attacks Using Mindfulness Techniques. *Journal of Management Information Systems*, *34*(2), 597–626. https://doi.org/10.1080/07421222.2017.1334499

[12] Kayode-Ajala, O. (2022). Applying Machine Learning Algorithms for Detecting Phishing Websites: Applications of SVM, KNN, Decision Trees, and Random Forests. *International Journal of Information and Cybersecurity*, *6*(1), 43-61.

[13] L. Anh, T. Nguyen, and H. K. Nguyen (2015). "Developing an efficient fuzzy model for phishing identification." In Control Conference (ASCC), 2015 10th Asian, pp. 1-6. IEEE.

[14] Li, Q., Cheng, M., Wang, J., & Sun, B. (2020). LSTM based phishing detection for big email data. *IEEE transactions on big data*, *8*(1), 278-288.

[15] Marivate, V., & Moiloa, P. (2016, November). Catching crime: Detection of public safety incidents using social media. In *2016 Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference (PRASA-RobMech)* (pp. 1-5). IEEE.

[16] M. S. Gerber, (2014) ‘‘Predicting crime using Twitter and kernel density estimation,’’ Decis. Support Syst., vol. 61, pp. 115–125.

[17] Naqvi, B., Perova, K., Farooq, A., Makhdoom, I., Oyedeji, S., & Porras, J. (2023). Mitigation strategies against the phishing attacks: A systematic literature review. *Computers & Security*, *132*, 103387. https://doi.org/10.1016/j.cose.2023.103387

[18] N.Zainuddin, A. Selamat, and R. Ibrahim (2016). ‘‘Improving Twitter AspectBased sentiment analysis using hybrid approach,’’ in Intelligent Information and Database Systems, vol. 9621, N. T. Nguyen, B. Trawinski, H. Fujita, and T.-P. Hong, Eds. Berlin, Germany: Springer, pp. 151–160

[19] Tam, S., & ÖzgürTanrıöver, Ö. (2023). Multimodal Deep Learning Crime Prediction Using Crime and Tweets. *IEEE Access*.

[20] Rathee, D., & Mann, S. (2022). Detection of E-Mail Phishing Attacks – using Machine Learning and Deep Learning. *International Journal of Computer Applications*, *183*(47), 1–7. https://doi.org/10.5120/ijca2022921868

[21] Sahingoz, O. K., Buber, E., Demir, O., & Diri, B. (2019). Machine learning based phishing detection from URLs. *Expert Systems with Applications*, *117*, 345-357.

[22] Şentürk, Ş., Yerli, E., & Soğukpınar, İ. (2017, October). Email phishing detection and prevention by using data mining techniques. In *2017 International Conference on Computer Science and Engineering (UBMK)* (pp. 707-712). IEEE.

[23] Shaukat, M. W., Amin, R., Muslam, M. M. A., Alshehri, A. H., & Xie, J. (2023). A hybrid approach for alluring ads phishing attack detection using machine learning. *Sensors*, *23*(19), 8070.

[24] S. K. Shinde (2014). "Detection of Phishing Websites Using Data Mining Techniques." International Journal of Engineering Research and Technology. Vol. 2. No. 12 (December-2013). ESRSA Publications.

[25] Singh, C. (2020, March). Phishing website detection based on machine learning: A survey. In *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)* (pp. 398-404). IEEE.

[26] Suzuki, Y. E., & Monroy, S. A. S. (2022). Prevention and mitigation measures against phishing emails: A sequential schema model. *Security Journal*, *35*(4), 1162–1182. https://doi.org/10.1057/s41284-021-00318-x

[27] S. Wedyan and F. Wedyan, (2013). "An Associative Classification Data Mining Approach for Detecting Phishing Websites." Journal of Emerging Trends in Computing and Information Sciences 4, no. 12.

[28] Vo, T., Sharma, R., Kumar, R., Son, L. H., Pham, B. T., Tien Bui, D., Priyadarshini, I., Sarkar, M., & Le, T. (2020). Crime rate detection using social media of different crime locations and Twitter part-of-speech tagger with Brown clustering. *Journal of Intelligent & Fuzzy Systems*, *38*(4), 4287–4299. https://doi.org/10.3233/JIFS-190870

[29] Wang, M. S. Gerber, and D. E. Brown. (2012). ‘‘Automatic crime prediction using events extracted from Twitter posts,’’ in Social Computing, Behavioral-Cultural Modeling and Prediction. Berlin, Germany: Springer, pp. 231–238.

[30] X. Chen, Y. Cho, and S. Y. Jang (2015).‘‘Crime prediction using Twitter sentiment and weather,’’ in Proc. Syst. Inf. Eng. Design Symp., Apr. pp. 63–68

[31] Zhu, E., Ju, Y., Chen, Z., Liu, F., & Fang, X. (2020). DTOF-ANN: an artificial neural network phishing detection model based on decision tree and optimal features. *Applied Soft Computing*, *95*, 106505.