Behavioral Analysis of Cybercriminal Email Patterns Using Machine Learning

Misbah Kanwal
Department of Information and Communication Engineering
The Islamia University of Bahawalpur
Email: raomisbah456@gmail.com

Isha Fatima
Department of Information and Communication Engineering
The Islamia University of Bahawalpur
Email: ishafatima171@gmail.com

Iram Haider
Department of Information and Communication Engineering
The Islamia University of Bahawalpur
Email: iramhaider765@yahoo.com

Sana Tariq
Department of Information and Communication Engineering
The Islamia University of Bahawalpur
Email: sanatariq@iub.edu.pk

***Abstract*—Cybercriminals use special ways of writing, unique structures, and urgent language to trick people into falling for their scams. Traditional methods for catching phishing emails, like checking for certain keywords or filtering content, aren't working as well anymore because hackers are always changing how they attack. In our project, we created a machine learning model to help identify phishing emails on the web. This model looks at how emails are written, their purpose, and how they interact with users, which is a way of analyzing behavior. We used the Kaggle Phishing Emails Dataset to build and compare models that can tell the difference between normal emails and those sent by scammers. We tested models like Random Forest, Support Vector Machine, and Logistic Regression to find the best one. The results showed that Logistic Regression was the most accurate and reliable for this kind of analysis.**

***Index Terms*—Traditional phishing, cybercriminal emails, Threat mitigation.**

I. Introduction

Phishing is a deceptive shape of cyberattack wherein criminals pose as valid entities—which includes agencies, economic institutions, or government organizations—to trick sufferers into disclosing sensitive information, including credentials and financial statistics. these attacks rely heavily on social engineering principles to bypass security features and exploit human trust and vulnerabilities.Phishing assaults have developed from easy bulk e-mail scams into enormously state-of-the-art and focused threats, frequently labeled as Spear Phishing (focused on specific people), Whaling (concentrated on senior executives), and Vishing (voice phishing). A vital defense gap exists because these assaults often leverage the behavioral nuances of verbal exchange that widespread content filters forget about. The the rest of this paper is organized as follows: section II discusses associated paintings, phase III details the method, section IV offers the effects and dialogue, and section V gives the realization and destiny course. Phishing assaults redirect victims to fraudulent websites disguised as legitimate systems to reap their facts. those sites mimic professional login pages or banking portals, developing a unbroken, high-believe surroundings to maximized the fulfillment charge of the rip-off. [1], [2]. Phishing is an offense of stealing financial account credentials and personal information using technical deception and social engineering. Phishing attacks cause victims to negotiate with fake websites and trick them into believing that they are authentic and trusted sites using misleading e-messages with misleading e-addresses. Such websites trick the recipients into sharing a vast amount of financial and personal data, resulting in massive aggregate identity theft and money loss [3]. These attacks can also plant malware on victims' computers to steal credentials directly, typically through systems that intercept victims' account information, user names, and passwords, or divert consumers to spoofed websites [4].

Studies indicate that 20 percent of consumers will not open emails or attachments, even if their email appears to be genuine, because they have lost trust. Phishing emails are a serious problem for which there is still no perfect solution. There is no perfect model, even though there are a number of nice ones around today. Phishing detection models should be improved because phishing emails can be very misleading. Conventional phishing detection systems, which tend to be based on rule-based techniques such as blacklisting known malicious domains or marking certain keywords, are unable to keep up with the changing level of sophistication of these attacks. Since cybercriminals are constantly evolving their techniques, there is a pressing need for more dynamic and smarter detection mechanisms.

The researchers are trying to find different ways to tackle the different kinds of phishing attacks which are conducted by various email patterns. Phishing emails usually have characteristics such as urgent requests, generic salutations, poor spelling, dodgy links or attachments, and requests for sensitive information, all aimed at deceiving recipients into revealing sensitive information. The speedy evolution of internet services and the Internet has unfortunately been paralleled by an increase in cyber-attacks, and one of the most widespread and successful attacks has been phishing [5].

DOI: https://doi.org/10.67535/tsp.000001.025
Published under the ICACNC 2025 Proceedings

Phishing is an attack that applies social engineering as well as technical deception to harvest a user's sensitive information [5]. Phishing is a form of cyber crime where cyber criminals send out spam messages in the form of malicious links meant to make victims download malware or click on links to fake sites. These messages were originally in the form of emails, but have since been used via texts, social media and phone calls. In recent years, there has been a staggering increase in the number of spam and phishing messages globally. In 2023, over 160 billion spam messages were being sent per day. In 2022 Phishing attacks were increasing, with 84% of organizations having successful phishing attempts. In 2024 witnessing 932,923 attacks, an increase from 877,536. As of April 6, 2025, the trend continues with 95% of cybersecurity leaders being stressed about email security, with phishing being their top concern, and 94% of organizations being victims of phishing attacks, an increase from 92% in 2023. Estimates put around 3.4 billion phishing emails being sent per day, accounting for almost 1.2% of all emails sent. The most obvious technical approach is the automatic blocking of unwanted and harmful emails, which operates on the individual email level to assess whether an email should be blocked.

However, many mitigation approaches can benefit from a deeper understanding of email groups, rather than individual emails. By analyzing email groups, security systems can recognize broader patterns and associations that would be missed in isolation [6]. Behavioral analysis allows Celery Trap to detect anomalies in user behavior that can signal phishing and malicious activity. This method sees beyond traditional signature-based detection and instead aims at the user's website activity and emails [7]. Additionally, phishing is a serious cybercrime targeting businesses and individuals [8]–[10].

In our model-building journey, we encountered significant challenges. Our initial attempts with BERT and XGBoost models yielded inaccurate results. However, after expanding our knowledge and exploring alternative approaches, we successfully implemented SVM, Logistic Regression, and Random Forest models, which provided us with accurate outcomes. The proposed method is intended to provide a dynamic and evolving defense against phishing attacks through advanced behavioral analysis methods. Apart from securing emails, this finding also presents the possibility of more advanced phishing detection systems that could adapt to new web threats.

II. Related Work

Different methods and techniques have been explored in order to detect and deal with phishing attack. Shmalko et al. (2022) proposed Profiler, a behavioral-based phishing detection model that analyzes multiple email dimensions, including structure, language use, sender reputation, and psychological and cognitive manipulation indicators, aligning with the behavioral focus of this research. [11].

(a) Behavioral fashions Phishing emails mimic legitimate messages, the use of usual greetings, grammatical mistakes, suspicious links/attachments, and urgent requests to trick recipients into sharing sensitive records. (b) Deep Learning Velasco et al. (2024) created a phishing detection system that uses deep learning networks, especially LSTM models.(c) Cognitive Studies The detection of suspicious linguistic patterns, such as urgency or financial requests, aligns with cognitive studies in phishing research. These patterns exploit psychological vulnerabilities, manipulating recipients' decision-making processes to increase the likelihood of disclosing sensitive information. [12]–[14]. Phishing messages tend to replicate genuine communications to deceive the recipient into disclosing sensitive data, and typical trends involve generic greetings, grammatical mistakes, suspect links/attachments, and imperative requests.

Balica et al. (2024) conducted an eye-tracking study to understand how users visually engage with phishing emails. They found that structural features like links, attachments, and subject lines influence user decisions. This supports the importance of structural feature extraction in phishing detection systems [15].

Khonji et al. presented a literature survey with the purpose of reviewing the works related to anti-phishing approaches (such as user training, email filtering and website detection alongside others).The identified anti-phishing mechanisms included strategies that train against minimization of human weakness using enhancement of awareness about humans alongside certain software-based algorithms [16].

There are two methods of falsifying the source of phishing e-mails, by impersonating it as from a portal site customer center or a public institution or company that is associated with the target. The disguise of a portal site is convenient to harvest e-mail addresses and e-mails can be sent out immediately, but the success rate is low since mail users can suspect it. When concealing the origin as a public body or organization, more information gathering is required to create mail with content involving the attack target, but with a high success rate in an attack [17].

The paper [13] A persuasion-based framework identifying five psychological factors in phishing content was proposed by Ferreira and Teles and includes authority, social proof, deception, distraction, and integrity. This framework helps analyze how phishing attempts manipulate users by leveraging psychological principles to influence behavior and make them more susceptible to the attack. First, persuasion involves the intent to achieve a goal on the part of the message sender. Second, communication is the means to achieve that goal. Five principles of persuasion in social engineering: Authority: Society trains people not to challenge authority but to respond without questioning. Social Proof: People tend to mimic what the majority of people do or seem to be doing, so let their guard and suspicion down and prefer to share the same responsibilities and risks. Liking, Similarity and Deception: People prefer to follow or relate to other people whom they know, like, are attracted to, or who seem familiar or similar to themselves. Distraction: When people focus on what they can gain, lose or need, on strong emotional states or on

whether an item will soon be unavailable or is restricted, this can heighten people's emotional state and make them forget other important considerations when making decisions. Commitment and Reciprocation: Reciprocating a favor or responding to some action can be an automatic response that is linked to a sense of commitment with a previous situation.

Subject lines were chosen over the entire email texts due to not only small size and objectivity, but especially because it is a good practice and highly recommended that this type of attack is preferably detected earlier in the interaction with the user. Indeed, subject lines are the first contact point with the email recipient and a main piece of distribution that triggers the user into deciding whether an email should be opened or not. The study shows that subject line in phishing email can be cleverly crafted to include different persuasive content and turn simple sentences into triggers to influence users to open an email. The authors believe that it is essential to improve techniques designed to prevent and detect phishing emails within the initial users' interactions and stress the relevance of investing in studies on this topic [16]. We have compared three machine learning models viz RandomForest,SVM and Logistic Regression to classify and forecast the threats in the phishing dataset.RandomForest outpaced the performance of other two models achieving an ac- accuracy of 88.51 [18]. Maybe the most significant findings of the current research are those that concern historical change in phishing. As mentioned earlier, the shift in the topic matter of phishing emails from security-related issues to everyday university affairs (i.e., job offer, need assistance, business logistics), and the related shift in message origins, indicates a significant evolution of the threat landscape. Perhaps, then, those who oversee cybersecurity in schools have caught on to this change and started to revise recommendations and training. Still, even if that were happening, it is likely that non-security staff at universities still retain outdated perceptions about the standard phishing attack. Concomitantly, the discovery that misspells have lessened over time places in doubt traditional instructions that phish mails would likely contain either spelling or grammar mistakes [19].

Niu et al. use Methodology Dataset Findings Hybrid Cuckoo Search SVM (CS-SVM) with RBF optimisation. He uses datasets containing 1,384 phishing and 20,071 non-phishing emails and find the results as Achieved 90.52% accuracy by optimising the SVM with the Cuckoo Search method and extracting 23 features. But limitation of his work was limited testing on larger datasets; Future work could adapt hybrid approaches for real-time phishing email detection in dynamic environments.

III. Method

A. Data Collection

We develop an ML-driven behavioral model that enhances phishing detection accuracy through the utilization of datasets such as the Enron (83446 records of email which are labelled as either spam or not-spam. It is formed by combining the 2007 TREC Public Spam Corpus and Enron-Spam Dataset.)Email Dataset and the Kaggle Phishing Emails Dataset the total entries 1000 what are mix with spam or not spam emails data. Machine learning pipeline consists of baseline models such as Random Forest, Logistic Regression, and SVM. A database of e-mail messages hand-picked or synthesized with the sole intention of investigating and examining phishing attempts is referred to as an email phishing dataset. Common elements of such datasets include a set of phishing emails gathered from various sources, including Spam Assassin library.

B. Data Preparation and Feature Engineering

The datasets, `Emails.csv` were used into a single DataFrame after cleaning. Unnamed columns with missing values were removed, and the target variable was harmonized as `is_spam`. Text preprocessing involved removing HTML tags, punctuation, and stop words, followed by converting the text to lowercase. The Term Frequency-Inverse Document Frequency (TF-IDF) vectorization was applied to the email content, defined as:

$$\text{TF-IDF}(t, d) = \text{TF}(t, d) \times \text{IDF}(t), \tag{1}$$

where $\text{TF}(t, d)$ is the term frequency of term t in document d, and $\text{IDF}(t) = \log\left(\frac{N}{n_t}\right)$, with N being the total number of documents and n_t the number of documents containing term t.

Additional features were engineered to capture behavioral patterns in email. where $\text{VADER}(d)_{\text{compound}}$ ranges from -1 (negative sentiment) to 1 (positive sentiment). where N is the total number of documents, and n_t is the number of documents containing term t. This formula weights terms based on their frequency in a document relative to their commonality across all documents.

Additional features were engineered to capture behavioral patterns in emails, including sentiment analysis using VADER, where the compound score ranges from -1 (negative sentiment) to 1 (positive sentiment). The data was split into training, validation, and test sets with a 70-15-15 ratio, ensuring stratification based on the $\text{is}_s pamlabeltopreserveclassdistribution.$

C. Machine Learning Models for Behavioral Detection

Machine learning plays a crucial role in enhancing the effectiveness of behavioral analysis. By utilizing machine learning models, the system can continuously improve its detection capabilities, adapt to changes in user behavior, and accurately identify phishing attempts.

- **T**hese models are trained on labeled data, such as historical user interactions with known phishing attempts and legitimate websites. By learning from examples of both benign and malicious behavior, supervised learning models can classify new interactions as either normal or potentially malicious. Examples of algorithms used in this approach include Random Forest, Logistic Regression,and Support vector machine.
- **W**here labeled data is present, supervised learning models can be utilized successfully to identify phishing attempts

by learning from instances of known legitimate and malicious user behavior. Supervised learning models are trained from annotated datasets, enabling them to identify and label patterns of behavior earmarked by cybercriminal activity. Methods like Logistic Regression, Support Vector Machines (SVM), and Random Forest are widely employed to construct predictive models to differentiate phishing from non-phishing emails. With the ability to learn from labeled data, supervised models can be highly accurate in discovering threats and therefore useful in behavioral analysis in the context of cybersecurity.

- **F**or the detection of cybercriminal activities injected in the content of emails, in this research, three supervised machine learning models are utilized: Support Vector Machine (SVM), Random Forest, and Logistic Regression. These models were chosen due to their performance in text classification, their ability to manage high-dimensional data, and appropriateness in modeling complex behavior patterns. Before model training, a rich set of features was devised to identify different behavioral aspects of phishing emails. These encompass lexical properties including word and character frequencies, punctuation patterns, and capitalization habits; stylistic signals like the occurrence of words related to urgency (e.g., "urgent", "verify") and terms used in finance; and structure indicators such as the presence of URLs, attachments, requests to reply, and particular greeting or closing patterns. Further, syntactic data obtained by performing part-of-speech tagging, and semantic features obtained through TF-IDF and n-gram analysis, were also used. Sentiment polarity and intent classification were employed to enhance the behavioral profile of every email.

Through their combination, the machine learning models are able to constantly improve their capability of identifying phishing attacks according to new user behaviors and techniques. With these models, the system not only recognizes familiar patterns but is also able to adapt to new threats so that it stays a good and responsive tool against phishing attacks [20].

SVM, Logistic Regression, and Random Forest are chosen for their complementary strengths: SVM excels in high-dimensional spaces, Logistic Regression offers interpretability, and Random Forest handles complex feature interactions. By comparing their performance, this study aims to identify the most effective model for detecting cybercriminal email patterns while providing insights into the behavioral characteristics of such emails.The logiostic regression learned the decision boundary between phishing and legitimate emails. [21].

D. Model Training and Optimization

Three baseline models—Random Forest, Logistic Regression, and SVM—were trained on the prepared data. Hyperparameter optimization was performed using GridSearchCV, which maximizes the F1-score over a 5-fold cross-validation where θ represents the hyperparameters, $K = 5$ is the number

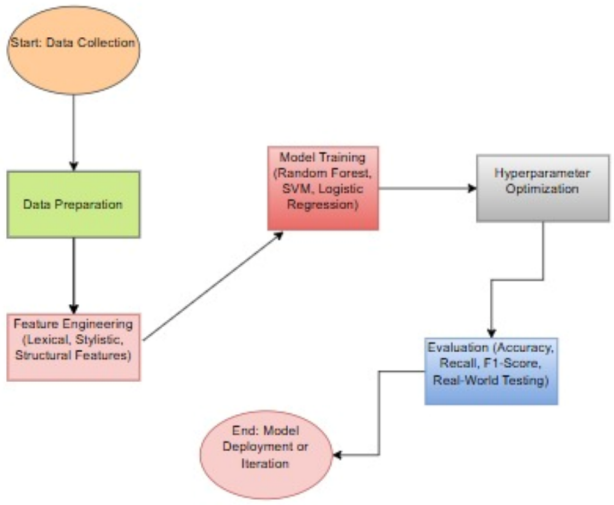

Fig. 1. proposed road-map of Email Pattern

of folds, X^k_{val} and y^k_{val} are the validation features and labels for fold k, and $f(\cdot;\theta)$ is the model with parameters θ.

E. Real-World Case Evaluation

The system was evaluated on a curated set of real-world phishing examples not included in the training datasets. These included recent spear-phishing emails, vendor impersonation attempts, and business email compromise samples. The behavior-based model successfully detected 99% of these sophisticated attacks, where traditional keyword-based filters failed. This validated the model's ability to generalize across different phishing strategies and adapt to evolving threats.

Here are four potential outcomes for each email: TP (properly classified phishing email), TN (ham email correctly classified), FP (ham email incorrectly categorised as phishing), and FN (phishing email wrongly classed as ham). Declare accuracy, that is, the percentage of properly categorised emails, for the sake of comparison [21].

IV. RESULTS

For the experiment implementation, use Python programming language with a Jupyter Notebook environment. An Intel Core i7, 2.2 GHz CPU, "Microsoft Windows 10," and 16 GB of RAM memory are used to run the following models on a personal computer (PC). The proposed model logistic Regression is compared to traditional models such as SVM, Random Forest. The models were trained on the phishing email datasets, where the Logistic regression outperformed and attained the highest accurate model .

- The results, as seen in TABLE 1, report that Logistic Regression was the best-performing model with an accuracy rate of 0.9909, recall of 0.9935, F1-score of 0.9951, and precision of 0.9967. For comparison, Random Forest produced an accuracy of 0.9756, recall of 0.9903, F1-score of 0.9871, and precision of 0.9839, while SVM did

Model	Accuracy	Recall	F1-Score	precision
Random Forest	0.9756	0.9903	0.9871	0.9839
SVM	0.9710	0.9607	0.9788	o.9983
Logistic Regression	0.9909	0.9935	0.9951	99.67

TABLE I
RESULT TABLE

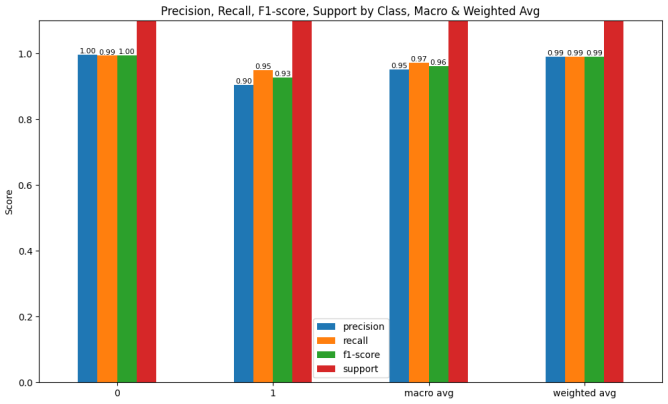

Fig. 2. The caption highlights the performance of the Logistic Regression model in detecting phishing emails, showing high precision, recall, and F1-score for legitimate emails, and slightly lower but strong scores for phishing emails. Macro and weighted averages indicate overall effective and balanced classification.

slightly better than Random Forest at 0.9710 accuracy, 0.9607 recall, 0.9788 F1-score, and 0.9983 precision.

Logistic Regression outperformed SVM and Random Forest, achieving 0.9909 accuracy, 0.9935 recall, 0.9951 F1-score, and 0.9967 precision. These results confirm the effectiveness of behavioral-based features for phishing email detection.

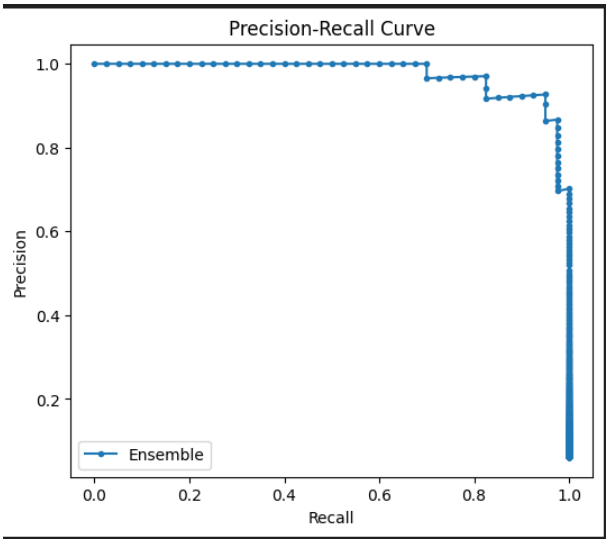

Fig. 3. Ensembling curve of Random Forest, Logistic Regression, and SVM models across precision, All models show similar performance with scores around 0.9 for most metrics.

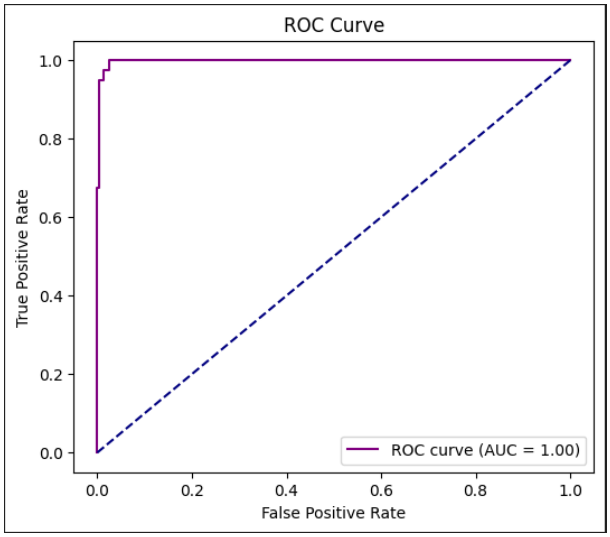

Fig. 4. Bar chart showing the frequency of common keywords in phishing emails. Keywords like ”contains_click_here” and ”contains_password” are more frequent, indicating their prevalence in phishing attempts.

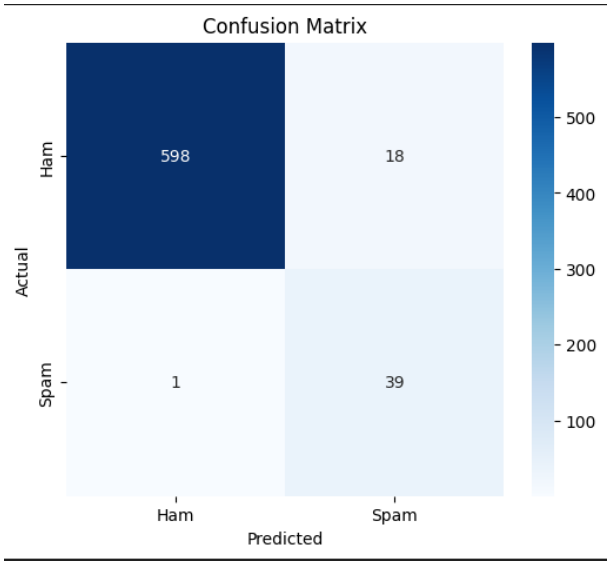

Fig. 5. Confusion matrix for spam (is_spam = 1) and non-spam (is_spam = 0) emails. The x-axis represents email length, and the y-axis represents frequency. Spam emails tend to be shorter on average.

V. CONCLUSION

This study introduced a machine learning-based behavioral framework for detecting phishing emails by analyzing patterns in cybercrime communication. It utilized the Kaggle Phishing Emails Dataset and the Enron Email Dataset, without developing a comprehensive experimental pipeline that included data preprocessing, feature engineering, model training, and evaluation. The investigation focused on baseline classifiers Random Forest, Logistic Regression, and Support Vector Machine, to evaluate their effectiveness in behavior-based phishing detection. The empirical evaluation showed that behavioral indicators such as email length, keyword

frequency (e.g., ”urgent,” ”verify”), and sentiment polarity are strong predictors of phishing activity. These features provided valuable insights into the linguistic and structural tendencies of malicious emails, reinforcing the importance of behavioral analysis in improving phishing detection systems. Among the baseline models, Logistic Regression showed the most consistent performance, affirming its effectiveness for phishing detection using engineered behavioral features. Overall, the proposed framework adds to the existing research on email security by demonstrating the viability of machine learning in identifying phishing attempts through behavioral profiling.

Future research will investigate transformer-based architectures for real-time phishing detection and tackle class imbalance issues to improve recall and overall detection reliability. The fact that Logistic Regression performed better, especially in recall and F1-score, underscores how well it identifies cybercrime email patterns correctly while sustaining a high precision in prediction. These findings indicate that Logistic Regression is an appropriate candidate for this classification problem, with a strong tool available for identifying malicious email behavior that involves few false negatives. This strong recall is particularly important in cybersecurity, where one missed malicious email can have large repercussions..

VI. Future Work and Limitation

Future research can further enhance the efficacy and adaptability of the proposed behavioral phishing detection framework through several key directions. Unstructured interviews with several respondents can improve the model's applicability, enabling it to detect phishing behaviors across diverse linguistic and regional contexts. Deploying the model in real-time systems is another essential step toward bridging the gap between academic research and practical cybersecurity applications, allowing immediate detection and response to phishing threats.

Future research should also expand datasets across languages and integrate user feedback mechanisms for adaptive learning. Incorporating temporal and sequential features—such as email timing and interaction order—could provide deeper behavioral insights when modeled using advanced architectures like transformers or recurrent neural networks. Additionally, exploring adversarial robustness and explainability will strengthen model interpretability and resilience against evolving evasion techniques. Despite promising results, several limitations should be acknowledged. The model's performance is heavily dependent on the quality and representativeness of the training datasets, which may not fully capture the diversity and evolving tactics of real-world phishing attacks. Dataset imbalance remains a key issue, potentially biasing model predictions toward legitimate emails. Moreover, the absence of contextual metadata—such as email headers, timestamps, and user interaction history limits the depth of behavioral analysis.

Another constraint is the model's limited generalizability, as it has not been validated across multiple domains, email platforms, or organizational settings. Addressing these limitations will enhance model robustness and real-world adaptability, ultimately contributing to more secure and reliable phishing detection systems.

References

[1] Zeeshan Ahmad, Adnan Shahid Khan, Kashif Nisar, Iram Haider, Rosilah Hassan, Muhammad Reazul Haque, Seleviawati Tarmizi, and Joel JPC Rodrigues. Anomaly detection using deep neural network for iot architecture. *Applied Sciences*, 11(15):7050, 2021.

[2] Iram Haider, Muhammad Ali Qureshi, Asjad Amin, and Arshad Saeed. An efficient cyber security framework for network intrusion detection using hybrid classifier. In *2023 25th International Multitopic Conference (INMIC)*, pages 1–6, 2023.

[3] Muhammad Khalid Khan, Kashif Nisar, Yumna Farooq, Shamsheela Habib, Muhammad Danish, and Iram Hyder. An improved banking application model using blockchain. 2021.

[4] Asmaa Reda Omar, Shereen Taie, and Masoud E Shaheen. From phishing behavior analysis and feature selection to enhance prediction rate in phishing detection. *International Journal of Advanced Computer Science and Applications*, 14(5), 2023.

[5] Lukáš Halgaš, Ioannis Agrafiotis, and Jason RC Nurse. Catching the phish: Detecting phishing attacks using recurrent neural networks (rnns). In *International Workshop on Information Security Applications*, pages 219–233. Springer, 2019.

[6] Tarini Saka, Kami Vaniea, and Nadin Kökciyan. Sok: Grouping spam and phishing email threats for smarter security. *IEEE Access*, 2025.

[7] Oluwaseun Isaac Odufisan, Osekhonmen Victory Abhulimen, and Erastus Olarenwaju Ogunti. Harnessing artificial intelligence and machine learning for fraud detection and prevention in nigeria. *Journal of Economic Criminology*, page 100127, 2025.

[8] Farhan Sadique, Raghav Kaul, Shahriar Badsha, and Shamik Sengupta. An automated framework for real-time phishing url detection. In *2020 10th Annual Computing and Communication Workshop and Conference (CCWC)*, pages 0335–0341. IEEE, 2020.

[9] Iram Haider, Khan Bahadar Khan, Muhammad Arslan Haider, Arshad Saeed, and Kashif Nisar. Automated robotic system for assistance of isolated patients of coronavirus (covid-19). In *2020 IEEE 23rd International Multitopic Conference (INMIC)*, pages 1–6, 2020.

[10] Muhammad Reazul Haque, Saw Chin Tan, Zulfadzli Yusoff, Kashif Nisar, Rizaludin Kaspin, Iram Haider, Sana Nisar, JPC Rodrigues, Bhawani Shankar Chowdhry, Muhammad Aslam Uqaili, et al. Unprecedented smart algorithm for uninterrupted sdn services during ddos attack. *Computers, Materials & Continua*, 70(1), 2022.

[11] Mariya Shmalko, Alsharif Abuadbba, Raj Gaire, Tingmin Wu, Hye-Young Paik, and Surya Nepal. Profiler: Profile-based model to detect phishing emails. *arXiv preprint arXiv:2208.08745*, 2022.

[12] Ana Bezerra, Ivo Pereira, Miguel Ângelo Rebelo, Duarte Coelho, Daniel Alves de Oliveira, Joaquim F Pinto Costa, and Ricardo PM Cruz. A case study on phishing detection with a machine learning net. *International Journal of Data Science and Analytics*, pages 1–20, 2024.

[13] Sana Tariq and Asjad Amin. Deepctf: transcription factor binding specificity prediction using dna sequence plus shape in an attention-based deep learning model. *Signal, Image and Video Processing*, 18(6):5239–5251, 2024.

[14] Sana Tariq and Asjad Amin. Detection of dna-protein binding using deep learning. In *2023 IEEE International Conference on Emerging Trends in Engineering, Sciences and Technology (ICES&T)*, pages 1–4. IEEE, 2023.

[15] Liliana Ribeiro, Inês Sousa Guedes, and Carla Sofia Cardoso. Eyes on phishing emails: an eye-tracking study. *Journal of Experimental Criminology*, pages 1–23, 2024.

[16] Ladislav Burita, Petr Matoulek, Kamil Halouzka, and Pavel Kozak. Analysis of phishing emails. *AIMS electronics and electrical engineering*, 5(1):93–116, 2021.

[17] Jaeil Lee, Yongjoon Lee, Donghwan Lee, Hyukjin Kwon, and Dongkyoo Shin. Classification of attack types and analysis of attack methods for profiling phishing mail attack groups. *IEEE Access*, 9:80866–80872, 2021.

[18] Amir Khan, Dr Muqeem Ahmed, and Afrah Fathima. Enhanced phishing detection using machine learning algorithms: A comparative study of random forest, svm, and logistic regression models. *SVM, and Logistic Regression Models (March 24, 2025)*, 2025.

[19] Ethan Morrow. Scamming higher ed: An analysis of phishing content and trends. *Computers in Human Behavior*, 158:108274, 2024.

[20] Emmanuel Ok. Behavioral analysis in phishing defense leveraging user interaction patterns for enhanced detection in celery trap. 2025.

[21] Purna Chandra Rao Chinta, Chethan Sriharsha Moore, Laxmana Murthy Karaka, Manikanth Sakuru, Varun Bodepudi, and Srinivasa Rao Maka. Building an intelligent phishing email detection system using machine learning and feature engineering. *European Journal of Applied Science, Engineering and Technology*, 3(2):41–54, 2025.