Minimizing Human Elements in Phishing Attacks by Integrating Administrative Controls to Phishing Detection Systems

Jonathan Owen Pang[1], Intan Farahana Binti Kamsin[2], Salmiah Binti Amin[3], Nur Khairunnisha Binti Zainal[4]
Asia Pacific University of Technology and Innovation, Technology Park Malaysia, Bukit Jalil, Kuala Lumpur, Malaysia.
[1]jon.op.academic@gmail.com, [2]intan.farahana@apu.edu.my, [3]salmiah@apu.edu.my,
[4]khairunnisha.zainal@apu.edu.my

Abstract – Phishing emails are a prevalent threat in the current cybersecurity landscape and industry-standard phishing detection system employed are considered insufficient in preventing phishing attacks as the detection system prioritizes on filtering and blocking phishing emails. This research aims to propose an enhanced phishing detection system addressing the limitation of the existing system deployed by integrating administrative countermeasures to reduce the number of successful phishing attacks. Regarding research methodology, this research will focus on gathering both quantitative and qualitative data by distributing surveys comprising of open-ended and close-ended questions and employing stratified sampling method to draw conclusions on the data collected. In conclusion, the proposed system highlights signs of phishing in an email which alleviates human error resulting in lower rate of successful phishing attacks. Future recommendation regarding the proposed system involves using an advanced Artificial Intelligence branch like Deep Learning and adding semantic analysis capability to the proposed system.

Index Terms – Machine Learning, Phishing Detection System, User Notification, Support Vector Machine (SVM)

1.0 Introduction

Phishing is a well-known social engineering attack that is used by cybercriminals to steal sensitive and confidential information from victims [5]. Within the constraints of emails, robust phishing detection system in the form of spam filtration system have been developed and deployed by popular email service providers like Google (Gmail) and Microsoft (Outlook). However, phishing attacks are generally mass distributed to potential victims which some may avoid the spam detection system. In addition, it is difficult to discern the legitimacy of a sophisticated spear-phishing email. Any phishing emails that successfully evaded the filtration system will be received by the recipient. The recipient in turn, may overlook signs of an email being phish due to human error like carelessness or the lack of phishing awareness training. Hence, by the sheer volume of sophisticated phishing emails sent, it is no surprise that many individuals and organizations were breach resulting in massive financial and reputational damage. For instance, [20] stated that the average cost to resolve a breach caused by phishing attack is around 14.8 million USD. Hence, this research aims to resolve the human aspect in phishing attack by proposing a system capable of detecting and highlighting signs of a malicious email which will be further elaborated in this report.

2.0 Literature Review

2.1 Domains

2.1.1 Overview on the Current Phishing Landscape

Phishing is a popular social engineering attack employed by cyber criminals, also known as phishers, to attain sensitive and confidential information such as Personally Identifiable Information (PII) and account credentials. According to [5], phishing is defined as a felony employing both social engineering and technical subterfuge to attain users' private information. Many technical and administrative countermeasures have been proposed to minimize the occurrence of successful phishing attacks. For instance, administrative countermeasure involves educating end users regarding risks and red flags found in phishing emails and technical countermeasure involves employing software capable of detecting common phishing features like presence of JavaScript in an email. However, these countermeasures are deemed inadequate as evidenced by the recent phishing landscape.

According to [8], Business Email Compromises (BEC) have increased in 2021 due to increased sophistication. In addition, Phishing-as-a-Service (PhaaS) are also gaining in popularity [8]. [28] stated that 85% of breaches detected in 2021 involved human elements and 61% of breaches involved credentials in their 2021 Data Breach Investigation Report (DBIR). The rate of phishing attack from 2019 to 2020 has increased by 220% according to [6].

These reports proves that phishing detection systems are not robust enough to detect all phishing emails. For instance, according to [2], phishing detection system using URL Features can be easily bypassed by using Generative Adversarial Networks (GAN). In this context, GAN is a machine learning framework where one neural network (Generator) generates "real" URL links and another neural network (Discriminator) attempt to discern whether the generated URL is real or fake. Thus, while

technical solution in powerful, it is still insufficient in filtering all phishing emails.

Regarding administrative countermeasures, the effect of phishing awareness training generally wears off as time passes. According to [22], the retention period of phishing awareness training is 4 to 5 months. The research made by [13] has also reflected the same results where retention of phishing awareness training is 5 months. In addition, personality traits, current state of emotions, stress levels, and general carelessness could be factors that lead to successful phishing attack regardless of phishing awareness retention [3], [13].

It is safe to conclude that both technical and administrative countermeasures are not absolute in preventing phishing attacks. Therefore, both technical and administrative countermeasures will be integrated while developing the proposed system to identify potential phishing email, pointing out signs of phishing, and leaving the categorization of spam or legitimate email to the user.

2.1.2 Selection of Machine Learning (ML) Algorithm

Machine Learning (ML) falls under the branch of Artificial Intelligence (AI) where data and algorithms are employed to imitate the human learning process [11]. ML can be subcategorized into Supervised ML, Unsupervised ML, Reinforcement ML, and Semi-Supervised ML [11], [24]. Within the context of this research, Supervised ML will be the focal point of research. Supervised ML algorithms are employed mainly for classification and regression. Classification refers to categorizing inputs or data according to features extracted, and regression refers to projecting predictions by interpreting relationships between dependent and independent variables. In this case, emails can either be classified as spam or legitimate, referred to as Binary Classification [12], [24].

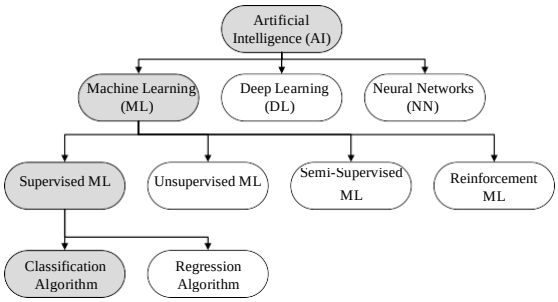

Figure 1 Branches of ML focusing on Classification Algorithm.

[21], [10], [18] succinctly stated the process of classifying the legitimacy of an email comprises of 5 steps:

1. Gathering of Data.
2. Pre-processing of Data.
3. Extraction of Features.
4. Classification of Data.
5. Analysis and Prediction of Email Legitimacy.

Hence, each major supervised ML algorithm will be reviewed according to its precision and accuracy in distinguishing the legitimacy of emails to select an optimal algorithm for the proposed system. Precision refers to the consistency in categorizing emails and accuracy refers to the rate of successful email categorization.

[19] compared multiple supervised classification ML algorithms including Support-Vector Machine (SVM), Naïve Bayes (NB), Random Forest (RF), and Decision Table (DT). The research concluded that SVM, NB, and RF consistently distinguished the legitimacy of emails with high precision and accuracy regardless of sample size.

[21] compared the accuracy and precision of SVM, RF, Logistic, NB, and VotedPreceptron. They concluded that SVM and RF have identical accuracy and precision, and are slightly better than NB.

[16] compared SVM, RF, NB, K-Nearest Neighbour (KNN), DT, and Adaboost. Similarly, both SVM and RF outperforms other ML algorithms.

Throughout the literature reviewed, RF and SVM consistently have the highest precision and accuracy compared to other ML algorithms. Hence, RF and SVM will be taken into consideration during the development of the proposed system. These supervised classifier ML algorithms will be employed to identify signs of phishing emails which will be used specifically to notify users to take precautions.

2.1.3 Selection of Features for Analysis

The process of distinguishing emails as phish or legitimate depends on the ML model derived using a training dataset and an appropriate ML algorithm [4]. According to [7], [14], [25], data must be processed and extracted in form of features before feeding into a supervised ML algorithm.

Initially, data are processed by removing stop words and missing data, stemming, lemmatizing, and tokenization of data [7], [14], [25] [29]. Subsequently, features from the processed data are extracted and split into two groups, a training dataset, and a testing dataset. The training dataset is fed into the ML algorithm to build a reliable ML model which is used for email prediction. Hence, the selection of features used as parameters for the ML algorithm is extremely crucial to building a reliable ML model.

[1] proposed a system focusing on classifying phishing emails using Random Forest ML algorithm which yielded an accuracy of 99.7% and a false-positive rate of 0.06%. The email features taken into consideration during this research are IP Addresses in URL links, disparities in Hyperlink references with link text, presence of JavaScript in email,

number of full stops in domain names, and keyword features.

[29] considered the number of domains in the emails, disparities in Hyperlink references and link text, number of dots in URLs in an email, URLs containing IP Addresses, Images as URLs, and keyword features when training their ML model to classify legitimacy of emails.

[14] evaluated word features, mismatched Hyperlinks references to its link text, visibility of link in plain text, presence of IP Address in URL, length of URL in terms of characters, number of dots in URLs, and invalid URL in the body of emails.

Previous research has conducted an in-depth investigation to identify notable features in emails that are used to distinguish the legitimacy of an email. Hence, this research will take the features specified into consideration when processing data and extracting features before feeding into the selected ML algorithm implemented in the proposed system.

2.2 Similar Systems

2.2.1 THEMIS

THEMIS is an email phishing detection model proposed by Fang et al. (2019) which employs a combination of multilevel embedding, enhanced Recurrent Convolutional Neural Network (RCNN), and an attention mechanism. [9] stated that an email comprises of two distinct components which are the email header, and the email body. The multilevel embedding component of THEMIS focuses on feature extraction on the semantics of words. [9] employed the Word2vec model to build a character-level and a word-level embedding model. In this case, word embedding is the process of representing the meaning of words in a vector. For instance, the word 'Queen. Subtracting the 'woman-ness' from 'Queen' and adding 'manly-ness' results in the word 'King' [30]. Figure 2 shows the way characters and words are represented in the THEMIS model.

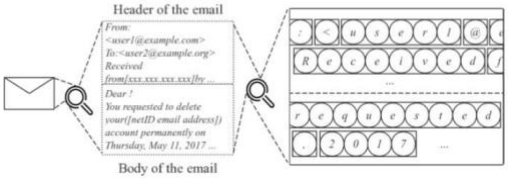

Figure 2 Representation of Characters and Words in THEMIS. Source: https://ieeexplore.ieee.org/document/8701426

The character and word features extracted are fed into the deep learning algorithm, Recurrent Convolutional Neural Network (RCNN). [9] enhanced the deep learning algorithm, RCNN proposed by [15] by integrating a Bidirectional Long-Short Term Memory (BI-LSTM) to overcome the issue faced by RCNN which is the inability to retain information for an extended period. An attention mechanism is deployed in THEMIS to elicit a stronger focus on certain aspects of an email's header and body.

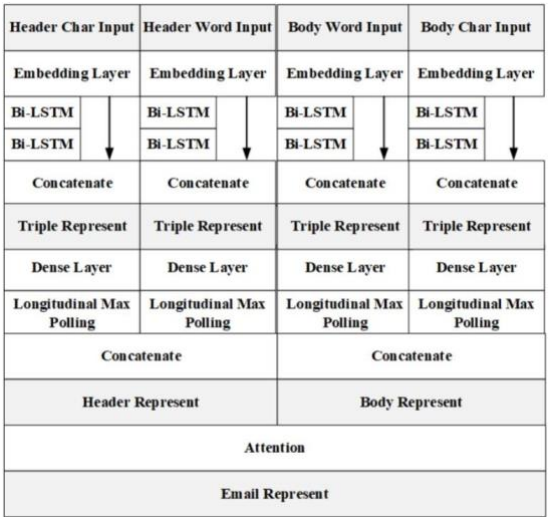

Figure 3 Overview on THEMIS model in classifying emails. Source: https://ieeexplore.ieee.org/document/8701426

The THEMIS model is developed by feeding large amount of test data comprising of both phish and legitimate emails. Subsequently, any test-validation email fed into the model will evaluated according to the probability of the email being phish or non-phish.

2.2.2 Phishing Detection using Random Forest, SVM, and Neural Network with Backpropagation

[26] proposed a system capable of distinguishing the legitimacy of a website using Artificial Intelligence. The researchers proposed the Machine Learning algorithm, Random Forest (RF), and Support Vector Machine (SVM). In addition, an advanced branch of Artificial Intelligence, Neural Network with Backpropagation was also proposed as these algorithms are used to classify a website as phish or legitimate. [26] evaluated the efficacy of each algorithm and concluded SVM is the most optimal algorithm which had the best accuracy. The proposed system was developed as a Chrome browser extension using the programming language, JavaScript. In addition, a major functionality of the system proposed by [26] is to notify users if a website is illegitimate or phish.

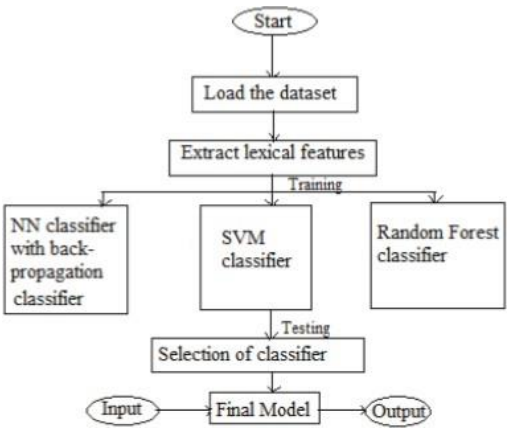

Figure 4 Workflow of the Browser-based Website Classification System.
Source: https://ieeexplore.ieee.org/document/9277256

Figure 4 shows the system workflow proposed by [26] which shows the process of selecting an optimal classifying algorithm and the system's input-output process.

2.2.3 PILFER

[10] proposed a system called PILFER which incorporates machine learning algorithms to classify both emails and websites as phish or legitimate. In this case, various features from both emails and websites are extracted such as IP-based URLs and the presence of JavaScript in emails. After extracting email and website features, [10] trained the Random Forest (RF) Machine Learning classifier algorithm using 10-fold cross-validation which is used to separate training data from test data. Subsequently, [10] stated that PILFER can be deployed in conjunction with an existing spam filtration system or as a standalone system.

While PILFER is proposed in 2007, several aspects of PILFER such as the features of email and website extracted can be used to reinforce the proposed system of this report.

2.3 Table of Comparison

As shown in Table 1, THEMIS focuses on classifying emails as phish or legitimate while the browser-based phishing detection system proposed by [26] emphasizes on website phishing detection. PILFER, on the other hand, focuses on classifying both emails and websites.

THEMIS employed an enhanced deep learning algorithm called Recurrent Convolutional Neural Network (RCNN) while both [26] and PILFER employ Machine Learning algorithm Support Vector Machine (SVM) and Random Forest (RF) respectively.

While [9] did not specify a medium of application, [26] stated that their system will be developed as a browser-based extension on Chrome using JavaScript. [10] stated that PILFER can be deployed either as a standalone system or be deployed in conjunction with an existing spam filtration system.

Lastly, the system proposed by [26] implemented a form of notification to users if a website is deemed as phish while PILFER and THEMIS did not include such functionalities.

2.4 Conclusion

As shown in Table 1, there are several gaps found in each of the system being compared. For instance, THEMIS and PILFER emphasizes on the technical aspect of phishing countermeasures. Both system/models emphasize on classifying emails as phish or not phish. However, technical countermeasures are not absolute as the current phishing landscape has indicated an upward trend in successful phishing attack as mentioned in the literature review section. This indicates that phishing emails can bypass filtration system in place and subsequently, managed to deceive users to disclosing confidential information.

Hence, there is a need to implement some form of administrative countermeasure on top of the technical countermeasure. In the system proposed by [26], their browser-based system incorporated an element of administrative countermeasure which was to notify users that the website being perused is a phishing website. However, this application is only applicable to websites and not emails.

Thus, the proposed system of this research will address the gaps found in the mentioned systems by [9]-[10], [26] by taking both technical and administrative countermeasure into consideration.

Table 1: Comparison of Important Features and Components of similar system

System Features	THEMIS	Browser-based Phishing Detection	PILFER
Researchers	[9]	[26]	[10]
Category of phish detected	Email	Website	Website and Email
Artificial Intelligence Category	Deep Learning	Machine Learning	Machine Learning
Learning Algorithm	Improved Recurrent Convolutional Neural Network (RCNN)	Support Vector Machine (SVM)	Random Forest (RF)
Medium of Application	Not Applicable	Browser: Chrome Extension	Standalone or in conjunction with existing spam filtration systems
Notify Users	No	Yes	No
Method of Notification	No	Popup alert notifying user website is suspected to be phish.	No

3.0 Problem Statement

Despite implementing sophisticated technical and administrative countermeasures to detect and filter out phishing emails, cases of people falling victim to phishing attacks are still prevalent. As previously mentioned in the literature review, the current phishing landscape suggests phishing activity and successful phishing attacks are increasing which indicates there is a gap in phishing countermeasures.

Regarding technical countermeasures, [2] proved phishing detection systems using URL features can be bypassed using Generative Adversarial Networks (GAN). Machine Learning approaches face an issue where legitimate emails are flagged as spam and are filtered out. Hence, technical solutions are just a means rather than an absolute solution in mitigating phishing attacks.

On the other hand, administrative countermeasures involving educating end-users through phishing awareness training also fall short in mitigating phishing attacks. Despite increased awareness of phishing, users are inherently susceptible to phishing due to human elements like the occasional carelessness, personality quirks, and emotions. For instance, volatile emotions like anxiety, fear, and stress impair one's cognitive function, affecting their susceptibility to phishing attacks. For example, [31] noted the emergence of Covid-19-themed phishing emails in their research. [23], [17] concluded that volatile emotions like stress and fear are linked to the Covid-19 pandemic. Thus, phishing emails exploit the public fear towards the pandemic by hinting at a shortage in Personal Protective Equipment (PPE) like masks resulting in users flocking to malicious websites masquerading as legitimate e-commerce websites.

Hence, an effective technical countermeasure must consider phishing email bypassing the filter and the human elements contributing to user susceptibility to clicking phish emails, such as carelessness, to minimize the chances of successful phishing attacks.

4.0 Research Aim

This research aims to propose an enhanced phishing detection system addressing the limitation of the existing system deployed by integrating administrative countermeasures to reduce the number of successful phishing attacks.

5.0 Research Objectives

- To evaluate various Machine Learning algorithm which can accurately and precisely classify an email as legitimate or phish.
- To implement an alert system highlighting signs of phishing and notifying users to take precaution.
- To research and pinpoint important features in emails which can be used to enhance the accuracy and precision of classifying emails as legitimate or phish.

6.0 Research Significance

As mentioned in Section 2.1.1, technical phishing countermeasures are not perfect in classifying and filtering phishing emails which leads to some malicious emails arriving to its recipients. Administrative countermeasures such as phishing

awareness training is susceptible to unavoidable factors like carelessness. By implementing the proposed system, the likelihood of users falling victim to phishing attack will be reduced. Any malicious emails bypassing the phishing detection system can be manually evaluated by the recipient with the guidance of the proposed system as signs of phishing and red flags are highlighted by the proposed system.

7.0 Research Methodology

7.1 Respondents Identification

The target respondents of this research are employees who constantly interacts with email service providers like Outlook and Gmail. The expected number of respondents is roughly 150 respondents who should come from a diverse background like age, job title, and years of experience. These factors are considered to improve the accuracy of conclusion drawn in the sampling process.

7.2 Sampling Method

This research employs the stratified sampling method to draw conclusions regarding an entire population from a subset of the population. Stratified sampling method is a complex probabilistic sampling technique which involves dividing a population into multiple sub-groups called strata in the process of stratification. Respondents in each stratum share the same characteristics. For example, the respondents in the research are divided according to their job titles. Subsequently, respondents in each stratum are chosen at random to represent the larger population.

Hence, stratified sampling is employed in this research because this sampling method ensure all strata are represented in the research. Comparatively, cluster sampling divides respondents into groups and several groups are selected randomly to act as representative of the overall population. Hence, population with influencing characteristics like job title may be overlooked.

7.3 Data Collection Method: Survey

The proposed system mitigates email phishing attacks by examining user emails and highlighting signs of phishing which may raise ethical issues regarding data collection and user privacy. Hence, it is important to collect information regarding user opinion on how the proposed system functions. Thus, the most appropriate data collection method is through surveys which will be distributed to respondents through emails. In addition, the overall cost of distributing surveys is lesser compared to interviews as the data collected are generally standardized and can be automated compared to interview where the interviewer must interpret the answers given by the interviewee.

As the research targets a wide demographic of respondents, questions in the survey will be structured in an understandable manner while avoiding any technical jargons. The survey will be comprised mainly of close-ended questions in the form of Likert Chart to collect respondent's opinion on the proposed system and open-ended questions to gather information regarding respondent's background and any suggestions made for the proposed system.

8.0 Overview of the Proposed System

8.1 Model Generation

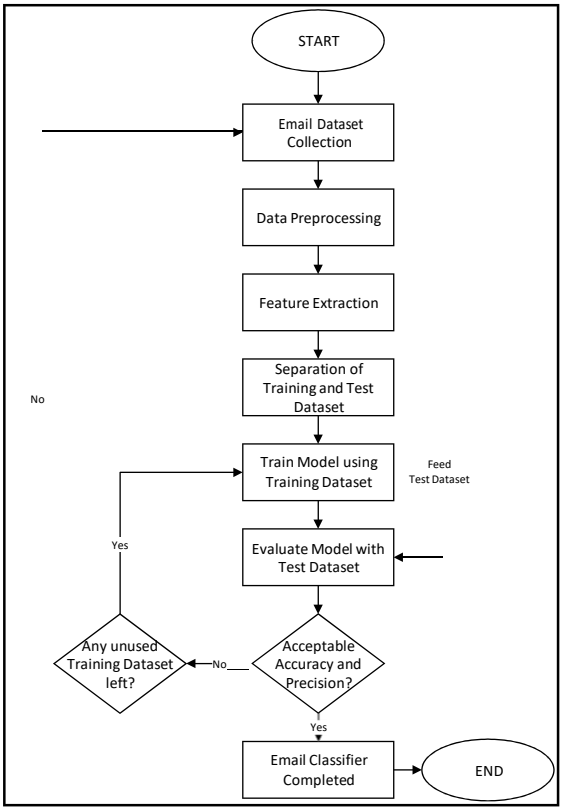

Figure 5 Process of developing a classifier model using SVM algorithm.

Figure 5 depicts the process involved in generating a model using Support Vector Machine (SVM) algorithm. The final output of this process is a robust model capable of classifying the legitimacy of emails with an acceptable rate of accuracy and precision.

8.2 Application of Proposed System

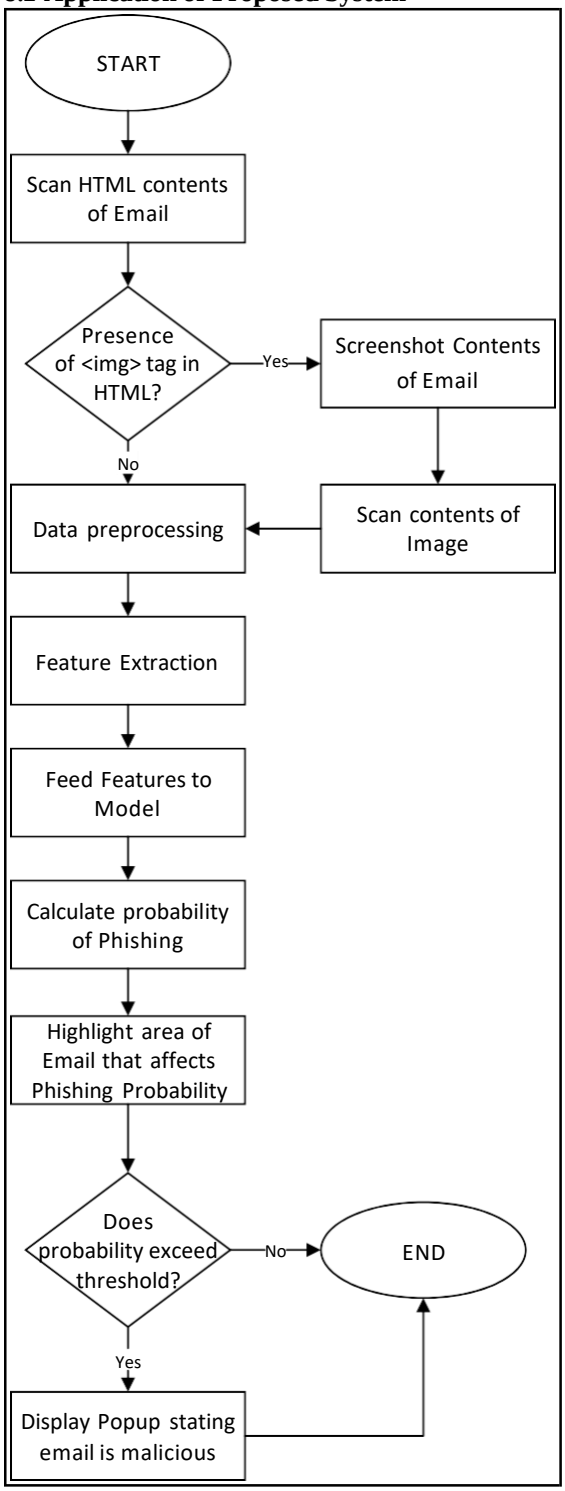

Figure 6 Process of classifying email as either legitimate or malicious.

Figure 6 demonstrates the process of classifying the legitimacy of an email after user input. Regardless of the email being legitimate or malicious, sections of an email contributing to the phishing probability will be highlighted.

9.0 Conclusion

This research proposes the integration of administrative countermeasures to address human elements in cybersecurity framework to enhance current phishing filtration mechanisms employed in mainstream email service providers. By highlighting crucial sections in an email and informing users how phishing emails are generally structured, human errors stemming from carelessness and negligence can be minimized. In addition, the proposed system can prolong the retention period of phishing awareness training as the system is constantly highlighting and notifying users regarding red flags found in malicious emails. Hence, the rate of successful phishing attacks decreases as users are more aware regarding signs of phishing emails.

10.0 Future Recommendations

Future recommendation regarding this proposed system involves employing a more robust and advanced Artificial Intelligence branches like Deep Learning. In addition, feature extraction could be improved by implementing a robust Artificial Intelligence model capable of performing semantic analysis as shown in the research made by [9].

11.0 Acknowledgement

I would like to offer my sincere appreciation and gratitude to Dr. Intan Farahana Binti Kamsin for her patience and guidance during my journey of writing this report. In addition, I would like to extend by gratitude to Asia Pacific University of Technology and Innovation for providing the opportunity to explore research methodologies.

12.0 References

[1] A. A. Akinyelu and A. O. Adewumi, "Classification of Phishing Email Using Random Forest Machine Learning Technique," *Journal of Applied Mathematics*, vol. 2014, pp. 1–6, Apr. 2014, doi: 10.1155/2014/425731.

[2] A. AlEroud and G. Karabatis, "Bypassing Detection of URL-based Phishing Attacks Using Generative Adversarial Deep Neural Networks," in *IWSPA '20: Proceedings of the Sixth International Workshop on Security and Privacy Analytics*, New Orleans, LA, USA, Mar. 2020, pp. 53–60. doi: 10.1145/3375708.3380315.

[3] Z. Alkhalil, C. Hewage, L. Nawaf, and I. Khan, "Phishing Attacks: A Recent Comprehensive Study and a New Anatomy," *Frontiers in Computer Science*, Mar. 2021, doi: 10.3389/fcomp.2021.563060.

[4] Amazon, "Training ML Models - Amazon Machine Learning," *Amazon*. https://docs.aws.amazon.com/machine-learning/latest/dg/training-ml-models.html (accessed Jan. 12, 2022).

[5] AWPG, "Phishing Activity Trends Report 2nd Quarter 2021," Anti-Phishing Working Group, Sep. 2021. Accessed: Jan. 10, 2022. [Online]. Available: https://docs.apwg.org/reports/apwg_trends_report_q2_2021.pdf

[6] CheckPoint, "CYBER SECURITY REPORT 2021," Checkpoint, 2021. Accessed: Jan. 10, 2022. [Online]. Available: https://www.checkpoint.com/downloads/resources/cyber-security-report-2021.pdf?mkt_tok=NzUwLURRSC01MjgAAAF7s-5iG39G8jZfk_sBjeqDJs5Egaj38DOLEb4lmGEwuqsP6C79eLeBRNqXeq7D0mJ6_Zmrej4GO7Ua2EsXDF2WymQ9ssrxbTCu7YHrvQOQ5KqCpE8

[7] E. G. Dada, J. S. Bassi, H. Chiroma, S. M. Abdulhamid, A. O. Adetunmbi, and O. E. Ajibuwa, "Machine learning for email spam filtering: review, approaches and open research problems," *Heliyon*, vol. 5, no. 6, May 2019, doi: 10.1016/j.heliyon.2019.e01802.

[8] ENISA, "ENISA Threat Landscape 2021," ENISA, Oct. 2021. Accessed: Jan. 10, 2022. [Online]. Available: https://www.enisa.europa.eu/publications/enisa-threat-landscape-2021

[9] Y. Fang, C. Zhang, C. Huang, L. Liu, and Y. Yang, "Phishing Email Detection Using Improved RCNN Model With Multilevel Vectors and Attention Mechanism," *IEEE Access*, vol. 7, pp. 56329–56340, Apr. 2019, doi: 10.1109/ACCESS.2019.2913705.

[10] I. Fette, N. M. Sadeh, and A. Tomasic, "Learning to Detect Phishing Emails," in *Learning to Detect Phishing Emails*, Banff, Alberta, Canada, Jan. 2007, pp. 649--656. Accessed: Jan. 12, 2022. [Online]. Available: https://www.researchgate.net/publication/221023704_Learning_to_Detect_Phishing_Emails

[11] IBM, "Machine Learning," *International Business Machine Corporation*, Jul. 15, 2020. https://www.ibm.com/my-en/cloud/learn/machine-learning (accessed Jan. 11, 2022).

[12] IBM, "Supervised Learning," *International Business Machines Corporation*, Aug. 19, 2020. https://www.ibm.com/cloud/learn/supervised-learning#toc-how-superv-A-QjXQz- (accessed Jan. 11, 2022).

[13] D. Jampen, G. Gür, T. Sutter, and B. Tellenbach, "Don't click: towards an effective anti-phishing training. A comparative literature review," *Human-centric Computing and Information Sciences*, vol. 10, no. 1, Aug. 2020, doi: 10.1186/s13673-020-00237-7.

[14] N. Kumar, S. Sonowal, and Nishant, "Email Spam Detection Using Machine Learning Algorithms," presented at the 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, Sep. 2020. Accessed: Feb. 06, 2022. [Online]. Available: https://www.researchgate.net/publication/344050184_Email_Spam_Detection_Using_Machine_Learning_Algorithms

[15] S. Lai, L. Xu, K. Liu, and J. Zhao, "Recurrent convolutional neural networks for text classification," in *AAAI'15: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, Austin Texas, Jan. 2015, p. Pg 2267–2273. Accessed: Feb. 06, 2022. [Online]. Available: https://dl.acm.org/doi/10.5555/2886521.2886636

[16] S. I. Manzoor and J. Singla, "A Comparative Analysis of Machine Learning Techniques for Spam Detection," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 8, no. 3, pp. 810–814, Jun. 2019, doi: 10.30534/ijatcse/2019/73832019.

[17] G. Mertens, L. Gerritsen, S. Duijndam, E. Salemink, and I. M. Engelhard, "Fear of the coronavirus (COVID-19): Predictors in an online study conducted in March 2020," *Journal of Anxiety Disorders*, vol. 74, p. 102258, Jun. 2020, doi: 10.1016/j.janxdis.2020.102258.

[18] I. Muhammad and Z. Yan, "SUPERVISED MACHINE LEARNING APPROACHES: A SURVEY," *International Journal on Soft Computing*, vol. 5, no. 3, Apr. 2015, doi: 10.21917/ijsc.2015.0133.

[19] F. Y. Osisanwo, J. E. T. Akinsola, O. Awodele, J. O. Hinmikaiye, O. Olakanmi, and J. Akinjobi, "Supervised Machine Learning Algorithms: Classification and Comparison," *International Journal of Computer Trends and Technology (IJCTT)*, vol. 48, pp. 128–138, Jun. 2017, doi: 10.14445/22312803/IJCTT-V48P126.

[20] Ponemon Institute, "The 2021 Cost of Phishing Study Sponsored by Proofpoint," Ponemon Institute, Traverse City, Michigan, United States, Jun. 2021. Accessed: Feb. 05, 2022. [Online]. Available: https://www.proofpoint.com/sites/default/files/analyst-reports/pfpt-us-ar-ponemon-2021-cost-of-phishing-study.pdf

[21] S. Rawal, B. Rawal, A. Shaheen, and S. Malik, "Phishing Detection in E-mails using Machine Learning," *International Journal of Applied Information Systems (IJAIS)*, vol. 12, no. 7, pp. 21–24, Oct. 2017, doi: 10.5120/ijais2017451713.

[22] B. Reinheimer *et al.*, "An investigation of phishing awareness and education over time: when and how to best remind users," in *SOUPS'20: Proceedings of the Sixteenth USENIX Conference on Usable Privacy and Security*, USENIX Association, 2560 Ninth St. Suite 215 Berkeley, CA, United States, Aug. 2020, pp. 259–284. Accessed: Jan. 10, 2022. [Online]. Available: https://www.usenix.org/system/files/soups2020-reinheimer_0.pdf

[23] A. J. Rodríguez-Hidalgo, Y. Pantaleón, I. Dios, and D. Falla, "Fear of COVID-19, Stress, and Anxiety in University Undergraduate Students: A Predictive Model for Depression," *Frontiers in Psychology*, vol. 11, Nov. 2020, doi: 10.3389/fpsyg.2020.591797.

[24] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN Computer Science*, vol. 2, no. 3, Mar. 2021, doi: 10.1007/s42979-021-00592-x.

[25] Z. B. Siddique, M. A. Khan, I. U. Din, A. Almogren, I. Mohiuddin, and S. Nazir, "Machine Learning-Based Detection of Spam Emails," *Scientific Programming*, vol. 2021, Dec. 2021, doi: 10.1155/2021/6508784.

[26] S. Sindhu, S. P. Patil, A. Sreevalsan, F. Rahman, and M. S. A. N, "Phishing Detection using Random Forest, SVM and Neural Network with Backpropagation," in *2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)*, Bengaluru, India, Dec. 2020, pp. 391–394. doi: 10.1109/ICSTCEE49637.2020.9277256.

[27] G. Sonowal, "Phishing Email Detection Based on Binary Search Feature Selection," *SN Computer Science*, vol. 1, no. 4, Jun. 2020, doi: 10.1007/s42979-020-00194-z.

[28] Verizon, "2021 Data Breach Investigations Report (DBIR)," Verizon, 2021. Accessed: Jan. 10, 2022. [Online]. Available: https://www.verizon.com/business/resources/reports/2021-data-breach-investigations-report.pdfx

[29] A. Yasin and A. Abuhasan, "An Intelligent Classification Model for Phishing Email Detection," *International Journal of Network Security & Its Applications*, vol. 8, no. 4, pp. 55–72, Aug. 2016, doi: 10.5121/ijnsa.2016.8405.

[30] Jason Brownlee, "What Are Word Embeddings for Text?," *Machine Learning Mastery*, Oct. 10, 2017. https://machinelearningmastery.com/what-are-word-embeddings/ (accessed Jan. 13, 2022).

[31] N. Akdemir and S. Yenal, "How Phishers Exploit the Coronavirus Pandemic: A Content Analysis of COVID-19 Themed Phishing Emails," *SAGE Open*, vol. 11, no. 3, Jul. 2021, doi: 10.1177/21582440211031879.