Department of Computing and Mathematics Computing and Digital Technology Postgraduate Programmes Terms of Reference Coversheet	
Student name:	Edafe Maxwell Damatie
University I.D:	23733083
Academic supervisor:	Dr. Amna Elenya
Project title:	Detecting and Preventing Phishing Attacks Using Machine Learning and Deep Learning Methods
Degree title:	MSc Cyber Security
Project unit code:	6G7V0007
Credit rating:	60
Start date:	30/05/2024
TOR date:	21/06/2024
Intended Submission date:	27/09/2024
Signature and date student:	Edafe Maxwell Damatie 26/06/2024
Signature and date external collaborator (if involved):	

1. Background and Rationale

Phishing attacks are a pervasive and significant threat in the digital age, targeting users' sensitive information through deceptive emails and websites. The increasing sophistication of these attacks necessitates advanced detection and prevention methods. Leveraging machine learning (ML) and deep learning (DL) techniques offers a promising solution to this challenge. Recent studies highlight the effectiveness of algorithms such as Random Forest, Decision Tree, and Neural Networks in identifying phishing attempts with high accuracy. This project aims to integrate the best performing algorithm into a user-facing solution, providing real-time protection against phishing attacks.

Phishing detection involves analyzing various features of emails and URLs to distinguish malicious content from legitimate communications. Studies have shown that deep learning methods, such as Deep Neural Networks (DNN), outperform traditional machine learning techniques by learning high-level features in an incremental manner, significantly improving detection accuracy, precision, and recall (Sumathi & Sujatha, 2019) (Ona et al., 2019). By combining feature selection and neural networks, we can develop robust models that effectively mitigate phishing threats (Martins de Souza et al., 2019).

2. Project Aims and Objectives

A. Aim

The aim of this project is to develop and implement machine learning and deep learning algorithms to detect and prevent phishing attacks. Additionally, I aim to create a user-friendly application that offers real-time phishing detection.

B. Objectives

- Conduct a comprehensive literature review on phishing detection methods.
- Design and implement algorithms such as Random Forest, Decision Tree, and Neural Networks.
- Evaluate the performance of the algorithms using standard benchmarks and datasets.
- Develop a user-friendly application integrating the implemented algorithms with the best performance and accuracy.
- Analyze the broader impact of the solution on privacy, security, and user trust.
- To critically assess and compare the effectiveness of various ML and DL algorithms in phishing detection.
- To address and evaluate the legal, ethical, professional, and social implications of deploying these technologies.

C. Research Questions

- What are the most effective features for detecting phishing emails and URLs?
- How do different machine learning and deep learning algorithms compare in terms of accuracy, precision, and recall in phishing detection?
- What are the ethical and legal considerations in deploying an automated phishing detection system?

3. Learning Outcome

- Detect phishing attacks using advanced techniques by implementing machine learning and deep learning methods such as Random Forest, Decision Tree, and Neural Networks.
- Customize and evaluate AI models specifically tailored for phishing detection to ensure high accuracy and efficiency.
- Process and analyze diverse data sources effectively by handling large volumes of data using appropriate tools and languages to create robust phishing detection systems.
- Develop AI solutions following best practices in software engineering, focusing on code reuse, modularity, testing, and comprehensive documentation.
- Integrate cybersecurity and AI principles by combining research and practical strategies to create innovative solutions for phishing detection and mitigation.

4. Project Description

A. Work to be Undertaken

i. Literature Review

Conduct a detailed review of existing literature on phishing detection methods, focusing on machine learning and deep learning approaches. Evaluate the strengths and weaknesses of different algorithms used in previous studies.

ii. Algorithm Design and Implementation

- **Random Forest and Decision Tree**: Implement these algorithms to classify emails and URLs as phishing or legitimate based on extracted features.

- **Neural Networks**: Develop a neural network model to improve detection accuracy through deep learning techniques.

iii. User-facing Application

Design and build an application that integrates the best performing algorithm, providing real-time phishing detection and prevention for users. Ensure the application follows best software development practices.

B. Legal, Ethical, and Professional Issues

Evaluate the ethical implications of using automated phishing detection systems, focusing on privacy concerns, data protection, and user consent. Address the potential legal issues related to data collection and analysis.

5. Evaluation Plan

A. Project Evaluation

The evaluation of this project will focus on determining if the project has met its aims and objectives:

- **Literature Review:** Check if the review covers the latest research on phishing detection methods.
- **Algorithm Implementation:** Verify that the Random Forest, Decision Tree, and Neural Networks are correctly designed and optimized for phishing detection.
- **Performance Measurement**: Assess algorithm effectiveness using benchmarks and like accuracy, precision, recall, and F1-score.
- **User-Friendly Application Development**: Ensure the application is user-friendly and integrates the best-performing algorithms.
- **Impact Analysis**: Evaluate the project's impact on privacy, security, and user trust.
- **Comparative Analysis** Compare the results with existing studies to determine the relative performance of the implemented solutions.

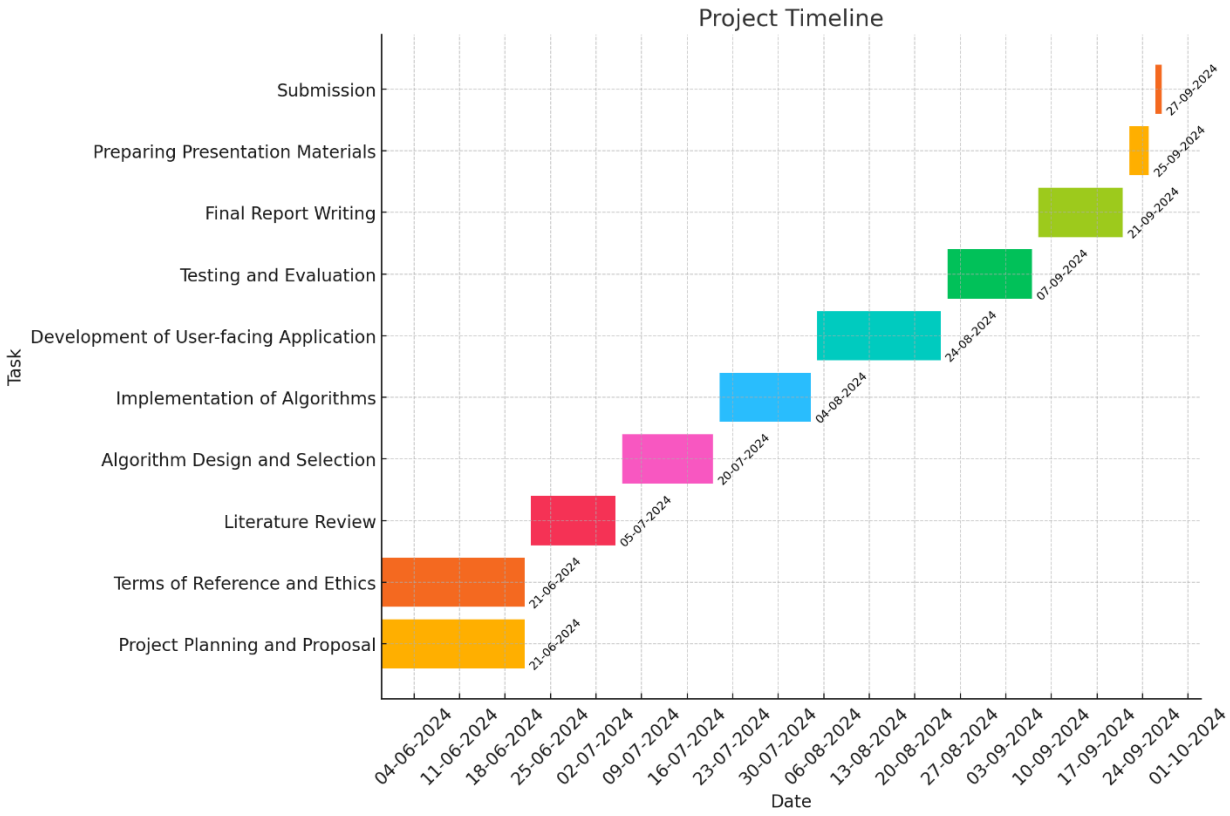
B. Product Evaluation

This evaluation will focus on determining if the final product meets its intended goals using following methods:

- **Algorithm Performance**: Measure effectiveness of the models using metrics like accuracy, precision, recall, and F1-score.
- **Usability Testing**: Assess user-friendliness through testing and feedback.
- **Real-Time Detection**: Verify the application provides accurate real-time phishing detection.
- **Privacy and Security Impact**: Ensure the application enhances security without compromising user privacy.
- **Integration and Functionality**: Check that algorithms are seamlessly integrated and all features function correctly.
- **Robustness and Scalability**: Test the application's performance with large data volumes and high usage.
- **Legal and Ethical Compliance**: Ensure adherence to relevant legal and ethical standards.

6. Activity Schedule

Task	Start Date	End Date	Duration (Days)
Project Planning and Proposal	30/05/2024	21/06/2024	22
Terms of Reference and Ethics	30/05/2024	21/06/2024	22
Literature Review	22/06/2024	05/07/2024	13
Algorithm Design and Selection	06/07/2024	20/07/2024	14
Implementation of Algorithms	21/07/2024	04/08/2024	14
Development of User-facing Application	05/08/2024	24/08/2024	19
Testing and Evaluation	25/08/2024	07/09/2024	13
Final Report Writing	08/09/2024	21/09/2024	13
Preparing Presentation Materials	22/09/2024	25/09/2024	3
Submission	26/09/2024	27/09/2024	1



7. References

- Rastogi, M., Chhetri, A., Singh, D., & V, G., 2021. Survey on Detection and Prevention of Phishing Websites using Machine Learning. 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), pp. 78-82. https://doi.org/10.1109/icacite51222.2021.9404714.
- Sumathi, K., & Sujatha, V., 2019. Deep Learning Based-Phishing Attack Detection. International Journal of Recent Technology and Engineering. https://doi.org/10.35940/ijrte.c6527.098319.
- Do, N., Selamat, A., Krejcar, O., Herrera-Viedma, E., & Fujita, H., 2022. Deep Learning for Phishing Detection: Taxonomy, Current Challenges and Future Directions. IEEE Access, 10, pp. 36429-36463. https://doi.org/10.1109/ACCESS.2022.3151903.
- Dewis, M., & Viana, T., 2022. Phish Responder: A Hybrid Machine Learning Approach to Detect Phishing and Spam Emails. Applied System Innovation. https://doi.org/10.3390/asi5040073.
- Tesfom, B., Belay, F., Daniel, S., Salem, R., & Otoum, S., 2023. Phishing Detection Using Deep Learning and Machine Learning Algorithms: Comparative Analysis. 2023 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech), pp. 0684-0689. https://doi.org/10.1109/DASC/PiCom/CBDCom/Cy59711.2023.10361457.

- Tang, L., & Mahmoud, Q., 2022. A Deep Learning-Based Framework for Phishing Website Detection. IEEE Access, 10, pp. 1509-1521. https://doi.org/10.1109/ACCESS.2021.3137636.
- Lakshmanarao, A., Rao, P., & Krishna, M., 2021. Phishing website detection using novel machine learning fusion approach. 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), pp. 1164-1169. https://doi.org/10.1109/ICAIS50930.2021.9395810.
- Kaushik, P., & Rathore, S., 2023. Deep Learning Multi-Agent Model for Phishing Cyber-attack Detection. International Journal on Recent and Innovation Trends in Computing and Communication. https://doi.org/10.17762/ijritcc.v11i9s.7674.
- Divakaran, D., & Oest, A., 2022. Phishing Detection Leveraging Machine Learning and Deep Learning: A Review. IEEE Security & Privacy, 20, pp. 86-95. https://doi.org/10.1109/MSEC.2022.3175225.
- Thakur, K., Ali, M., Obaidat, M., & Kamruzzaman, A., 2023. A Systematic Review on Deep-Learning-Based Phishing Email Detection. Electronics. https://doi.org/10.3390/electronics12214545.
- Aljabri, M., & Mirza, S., 2022. Phishing Attacks Detection using Machine Learning and Deep Learning Models. 2022 7th International Conference on Data Science and Machine Learning Applications (CDMA), pp. 175-180. https://doi.org/10.1109/CDMA54072.2022.00034.
- Vinayakumar, R., Soman, K., Poornachandran, P., Akarsh, S., & Elhoseny, M., 2019. Deep Learning Framework for Cyber Threat Situational Awareness Based on Email and URL Data Analysis. Advanced Sciences and Technologies for Security Applications. https://doi.org/10.1007/978-3-030-16837-7_6.
- Martins de Souza, C. H., Lemos, M. O., Dantas Silva, F. S., & Souza Alves, R. L. (2019). On detecting and mitigating phishing attacks through featureless machine learning techniques. Internet Technology Letters, 3(1), e135. https://doi.org/10.1002/itl2.135
- Oña, D., Zapata, L., Fuertes, W., Rodríguez, G., Benavides, E., & Toulkeridis, T. (2019, October). Phishing Attacks: Detecting and Preventing Infected E-mails Using Machine Learning Methods. In 2019 3rd Cyber Security in Networking Conference (CSNet) (pp. 161-163). IEEE. https://doi.org/10.1109/CSNet47905.2019.9108961.