DeepPhish-X: Multi-Modal Feature Engineering for Phishing Detection Using Hybrid Models of Computer Vision, Natural Language Processing, and Graph Neural Networks

Program Name:	**Masters of Science (Data Science)**
Subject Name:	**Project 1 (MCST1043)**
Student Name:	Cui ZhiWen
Metric Number:	MCS241040
Student Email & Phone:	cuizhiwen@graduate.utm.my
Project Title:	DeepPhish-X: Multi-Modal Feature Engineering for Phishing Detection Using Hybrid Models of Computer Vision, Natural Language Processing, and Graph Neural Networks
Supervisor 1:	
Supervisor 2 / Industry Advisor(if any):	

5. Conclusions

This study introduced **DeepPhish-X**, a phishing detection approach that integrates HTML DOM graph modeling with URL feature analysis using hybrid deep learning techniques.[1] By leveraging Graph Convolutional Networks (GCNs) from

Graph Neural Networks to model HTML DOM graphs, and using Convolutional Neural Networks (CNNs) from **Computer Vision** and Transformer Networks from **Natural Language Processing** to capture character and word sequence features from URLs, DeepPhish-X effectively combines these multi-modal features.[1] This approach addresses the limitations of traditional URL-based phishing detection methods, which often struggle to capture the full context of phishing attacks.[1]

DeepPhish-X demonstrated significant performance improvements, achieving a 7.03 percentage point increase in classification accuracy compared to existing state-of-the-art techniques.[1] The detailed evaluation, including confusion matrix analysis and ablation studies, highlighted the importance of integrating character-based URL features, word-based URL features, and HTML DOM graphs for effective phishing detection.[1] The results validated DeepPhish-X's superiority in accurately identifying phishing webpages, which is attributed to the complementary strengths of the URL multi-modal features and the diverse deep learning models used.[1] This research makes a significant technical contribution by being the first to combine URL features and HTML DOM graph features and by designing a neural network that effectively merges these complementary characteristics.[1]

Despite these improvements, some limitations were identified during the study.[1] DeepPhish-X relies on static features extracted from URLs and HTML DOM structures, which may be susceptible to obfuscation by evolving phishing tactics, potentially reducing the model's effectiveness over time.[1] Additionally, some benign URLs with complex structures were misclassified as phishing, indicating the need for additional contextual information (e.g., user behavior or dynamic content analysis) to enhance detection accuracy.[1] While the integration of multi-modal features improved detection rates, it also increased computational complexity, which could be a limitation in real-time applications where processing speed is crucial.[1]

Future research should focus on incorporating additional HTML features, such as DOM tag names and hyperlinks, to provide a more detailed representation of the webpage.[1] Integrating user behavior analysis and browser interaction patterns could offer deeper insights into

phishing detection by considering both static properties and dynamic interactions.[1] Additionally, leveraging advancements in adversarial learning to improve robustness against sophisticated phishing tactics is another promising direction.[1] These enhancements can further improve the efficacy of DeepPhish-X, making it more comprehensive and reliable.[1]

Works cited