Adaptive Phishing Detection on Messaging Platforms

This project focuses on the development of a phishing detection system specifically designed for real-time messaging platforms, such as Discord, which are increasingly becoming prime targets for phishing attacks. Traditional phishing detection systems have been primarily designed for email-based threats, where the language structure, content, and attack methods are more standardized. In contrast, real-time messaging platforms pose a unique challenge due to their dynamic and informal communication styles. This system leverages advanced Natural Language Processing (NLP) techniques and adaptive machine learning models that are capable of interpreting the nuances of informal language and rapidly changing conversation contexts. The core of the system is built upon deep learning algorithms, particularly Recurrent Neural Networks (RNNs), that analyze message content, detect potential phishing attempts, and classify the messages accordingly. Furthermore, the system dynamically learns from evolving phishing patterns, continually improving its detection capabilities in response to new threats. By using a feedback loop that allows real-time learning and adaptation, the system ensures that it stays updated and capable of providing real-time protection. This approach not only improves detection accuracy but also enhances the system's ability to adapt to sophisticated phishing strategies that may emerge on messaging platforms.

Evaluation of combined classification developed in the project:

SPAM DATASET RESULTS

Confusion Matrix

Actual / Predicted	Spam (1)	Ham (0)
Spam (1)	475	253
Ham (0)	45	4780

Fig. 5. The confusion matrix for the combined model on the spam dataset.

Classification Report

Class	Precision	Recall	F1-score	Support
Spam	0.91	0.64	0.75	747
Ham	0.95	0.99	0.97	4825
Micro avg	0.95	0.94	0.94	5572
Macro avg	0.93	0.81	0.86	5572
Weighted avg	0.94	0.94	0.94	5572

Fig. 6. The classification report for the combined model on the spam dataset.

PHISHING DATASET RESULTS

Confusion Matrix

Actual / Predicted	Phishing (1)	Ham (0)
Phishing (1)	2912	559
Ham (0)	99	5251

Fig. 7. The confusion matrix for the combined model on the phishing dataset.

Classification Report

Class	Precision	Recall	F1-score	Support
Phishing	0.97	0.66	0.79	4383
Ham	0.90	0.61	0.73	8641
Micro avg	0.93	0.63	0.75	13024
Macro avg	0.94	0.64	0.76	13024
Weighted avg	0.93	0.63	0.75	13024

Fig. 8. The classification report for the combined model on the phishing dataset.