Detecting Phishing Attacks Using Machine Learning in Email Systems

Loli Ray **Department** Information Technology

Background

Phishing attacks are among the most widespread types of cybercrime, establishing a significant method for thousands of organizations to steal passwords and other sensitive information from its user database. Phishing remains one of the most reported cybercrimes year after year with thousands of incidents happening each year (Perez, 2025) according to FBI's Internet Crime Complaint Center. Traditional email filtering systems rely heavily on rule-based detection, which can find it difficult to adapt to the constantly evolving and complex phishing techniques.

The increasing utilization of machine learning (ML) in cybersecurity, especially in identifying anomalies and patterns within large data sets is well documented (Sahoo et al., 2017). Machine learning algorithms can utilize features including email content, sender behavior, and message metadata for distinguishing messages as malicious or benign. However, most current systems are either too resource-intensive or don't adapt in real-time for everyday use.

As an Information Technology student focusing on secure computing and networks, I am interested in knowing how we can fine-tune lightweight machine learning models to enhance phishing detection with a hands-on approach. This project seeks to develop

and analyze an effective phishing detector, suitable for integration into common email infrastructures.

Methodology

This project will involve collecting and analyzing phishing and legitimate email datasets, as well as training and testing the machine learning models.

The first step will be to gather publicly used datasets like the Enron email dataset and phishing datasets from sources like Kaggle and UCI Machine Learning Repository. We will preprocess these datasets by cleaning up text, removing duplicates if necessary and extracting features based on each dataset such as URLs, keywords, sender domains, or email structure.

Then, I will implement several machine learning models in Python, including logistic regression, decision trees, and random forest. The models will be trained and tested using an 80/20 dataset split. The performance will be measured by accuracy, precision, recall, and F1-score.

Additionally, I will assess how these models perform on their ability to detect phishing emails, and evaluate which attributes seem most efficient in getting the correct classification. The last step will be developing a simple prototype system that shows how the model might mark suspicious emails in real time.

Participant Recruitment

Participants will be recruited from the University of Central Florida student population through campus listservs, student organizations, and flyers. Participants will complete a short task involving identifying phishing and legitimate emails. Each participant will receive a $10 digital Amazon gift card delivered via email upon completion of the study.

Anticipated Outcomes

This project is expected to produce the following research outcomes:

- Labeled dataset of phishing and legitimate emails with extracted features
- Comparative analysis of machine learning model performance
- Identification of the most effective features for phishing detection
- Formal research report summarizing findings
- Conference-style presentation suitable for undergraduate research showcases

Significance

This research contributes to improving cybersecurity practices by providing a more adaptive and efficient approach to phishing detection. It demonstrates how machine learning can enhance traditional email security systems with a wide range of intended audiences, such as IT and cybersecurity professionals/analysts, students interested in cybersecurity and machine learning, and organizations seeking cost effective security measures. There is also a wide variety of project benefits, including but not limited to identifying effective features for phishing detection, providing insight into practical

applications of machine learning in cybersecurity, demonstrating a prototype system for real-world use, and providing understanding of secure email systems and threat prevention.

References (MLA)

Al-Subaiey, Abdulla, et al. “Novel Interpretable and Robust Web-Based AI Platform for Phishing Email Detection.” *arXiv.Org*, 19 May 2024, https://arxiv.org/abs/2405.11619v1.

Mohammad, Rami and Lee McCluskey. "Phishing Websites." UCI Machine Learning Repository, 2012, https://doi.org/10.24432/C51W2X

Perez, Jose. *Federal Bureau of Investigation Internet Crime Report*. 2025, https://www.ic3.gov/AnnualReport/Reports.

Sahoo, Doyen, et al. “Malicious URL Detection Using Machine Learning: A Survey.” *arXiv.Org*, 25 Jan. 2017, https://arxiv.org/abs/1701.07179v3.

Timeline

Dates	**Tasks**
January 1, 2027 - January 31, 2027	Conduct literature review and identify datasets
February 1, 2027 - February 28, 2027	Collect data and begin preprocessing
March 1, 2027 - March 31, 2027	Train machine learning models
April 1, 2027 - April 30, 2027	Evaluate performance and finalize results and presentation

Budget

Item	**Units and Cost**	**Total**
Participant incentives	$10/person, 10 people	$100
NVivo license	Free through UCF IT services	$0
External hard drive	$60	$60
Programming tools (Python, Scikit-learn, NumPy, Pandas)	No fee, open source	$40
Cloud computing credits (Amazon Web Services EC2)	$150 (estimated usage)	$150
Total	**$310**	**$310**