CYB201/CIT251 Project Proposal

Detecting and Preventing Phishing Attacks Using "Machine Learning"

4ortified Team Members:

PM: Nicolas Rossi (nr71476n@pace.edu)

Eleon Annoor (ea93590n@pace.edu)

Ghardesh Dolcharran (gd66512n@pace.edu)

November 5th, 2025

Abstract:

Phishing attacks continue to be among the most common and destructive types of cybercrime, taking advantage of human psychology to get past technical defenses and breach user information. As a subset of malware-based social engineering, phishing deceives users into disclosing confidential information by impersonating reliable parties through emails, messages, or websites. Even with the use of conventional detection approaches like rule-based filters, heuristic analysis, and blacklists, attackers are always changing their tactics, making many of the defenses in place obsolete. By creating a machine learning based model that can reliably identify and stop phishing attempts through automated URL and email analysis, our research aims to address this discrepancy. Using Python along with libraries such as Scikit-learn and Pandas, this project will analyze the PhishTank dataset to train and test classification algorithms capable of distinguishing legitimate URLs from phishing links. To find predictive patterns, characteristics like domain structure, SSL certificate validity, URL length, and keyword presence will be extracted and examined. By comparing multiple supervised learning algorithms, the model aims to achieve a high detection accuracy while minimizing false positives. The anticipated result is a trained model that outperforms traditional filtering methods in terms of phishing identification accuracy. In addition to demonstrating the usefulness of machine learning in cybersecurity, the project's realistic strategy will highlight the importance of preventative measures against evolving phishing attacks. In the end, this effort strengthens human interaction, one of the most vulnerable points in the cybersecurity chain, making digital environments safer for both individuals and enterprises.

Introduction

Malware is one of the biggest threats in Cybersecurity, and there are so many different types of malware that can be utilized by a hacker. It's usually used to steal or recover confidential information, get through integral software/sites, and affect the availability of a source, which each of these breaches affects a different aspect of the CIA Triad. One of the most prominent malware attacks that we see today is Phishing attacks. It's a form of social engineering where the malware attempts to get access to valuable information by pretending to be some other entity. For example, if you ever receive a text message from some random number claiming to be a famous actor but in need of money, and they will pay you back after their next acting gig, that is a phishing attempt to gain money from you by using a famous person's identity. There are many other examples, including emails, phone calls, websites, links, etc. This type of malware is detrimental to the CIA Triad and causes much harm to people and their identities. We are motivated to help people protect their identities and their assets from the most common malware attack in an attempt to make the internet a safer place and one that people feel more comfortable using. 4ortified is determined to help decrease the risk that this malware creates for so many people and their assets by using "machine learning" to compile and test a phishing detection and prevention model. We expect to have a trained model that is capable of identifying phishing URLS or emails at a higher accuracy than standard filters through pattern recognition and classification.

Background

Over the past decade, phishing detection has been one of the most widely researched topics in cybersecurity due to the constant evolution of social engineering techniques. To help detect phishing websites, emails, and messages before people fall victim to them, numerous researchers have created and tested a variety of tactics. In the past, algorithmic criteria and static blacklists were used for phishing detection. However, these approaches often fail to keep up with the constantly changing patterns used by attackers. In order to offer more flexible and dynamic solutions, the cybersecurity sector started investigating machine learning techniques. In 2018, Mohammad et al. conducted research on “Intelligent Phishing Website Detection Using Machine Learning,” where they extracted a set of URL-based features and trained classification models to detect phishing websites with high accuracy. Following that, studies expanded on this strategy by adding new feature sets like SSL certificate checking, DNS information, and HTML tag analysis. In 2020, Patil and colleagues expanded this work using Random Forest and Decision Tree algorithms, achieving over 96% accuracy on phishing datasets. More recently, in 2023, University of New Haven researchers developed a hybrid model that combines URL-based data and natural language processing (NLP) to identify phishing emails based on both content and structure. Despite the remarkable accuracy rates attained by these studies, many models either need a lot of processing power or find it difficult to adjust when attackers change their strategies. By emphasizing both detection accuracy and practicality, our project expands upon these current frameworks. Using Python, SciKit-learn, and the PhishTank dataset, we intend to train a model that is optimized for practical usage and lightweight performance. We seek to develop a model that strikes a compromise between accuracy, speed, and adaptability for

routine phishing avoidance by comparing several algorithms and examining characteristics like URL length, domain age, and HTTPS presence.

Lab Design / Research Methodology

The lab for this project is designed to apply machine learning techniques to detect and prevent phishing attacks using real phishing data. The exercises involve building and testing a classification model that can differentiate between phishing and legitimate URLs. **Hardware and Software Tools:** macOS laptop or desktop computer, Python 3.8 or newer. Jupyter Notebook, Libraries: SciKit-learn, Pandas, NumPy, Matplotlib, Dataset: PhishTank verified phishing URLs combined with legitimate URLs from trusted sources

Technical Approach and Workflow:

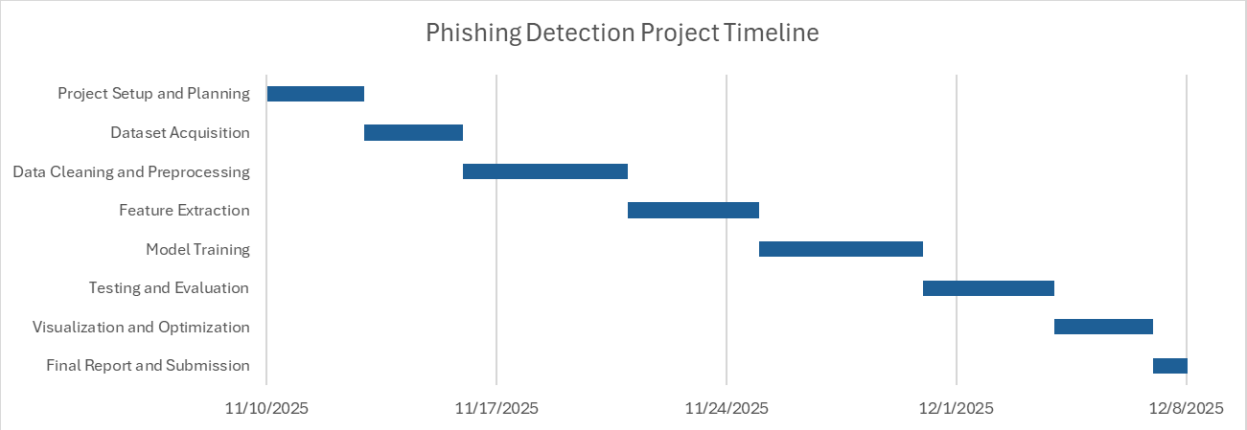
1. **Data Collection and Preprocessing:** Import phishing and legitimate URLs, clean the data, and remove duplicates or missing entries.
2. **Feature Extraction:** Derive measurable attributes such as URL length, presence of "@" symbols, HTTPS usage, and domain age.
3. **Model Training:** Use supervised machine learning algorithms (Decision Tree, Random Forest, Logistic Regression) in SciKit-learn to classify URLs.
4. **Evaluation:** Assess performance using accuracy, precision, recall, and F1 score metrics.
5. **Visualization:** Plot results using Matplotlib to compare algorithm performance.

This lab supports our investigation by allowing us to analyze how machine learning can identify phishing attempts based on URL characteristics. The workflow provides a hands-on understanding of feature selection, data preprocessing, and classification—core components of cybersecurity analytics.

Project Timeline

The project timeline spans from November 10th to December 8th, 2025, outlining each major phase from planning to final reporting. Tasks are arranged sequentially to ensure efficient workflow and progress tracking. Each milestone builds upon the previous one — starting with dataset acquisition and preprocessing, followed by feature extraction, model training,

evaluation, and optimization. The Gantt chart below visualizes this schedule, displaying the start and end dates and the task durations for the full project cycle.



References

1. Mohammad, R. M., Thabtah, F., & McCluskey, L. (2014). Predicting phishing websites based on self-structuring neural network. *Neural Computing and Applications, 25*(2), 443–458. https://doi.org/10.1007/s00521-013-1490-z

2. Patil, V., Thakkar, P., Shah, C., Bhat, T., & Godse, S. P. (2018). Detection and prevention of phishing websites using a machine learning approach. *2018 4th International Conference on Computing, Communication, Control and Automation (ICCUBEA)*, 1–5. IEEE. https://doi.org/10.1109/ICCUBEA.2018.8697412

3. Khan, S. A., Khan, W., & Hussain, A. (2021). Phishing attacks and websites classification using machine learning and multiple datasets: A comparative analysis. *arXiv preprint arXiv:2101.02552.* https://arxiv.org/abs/2101.02552

4. Gupta, S. D., Arachchilage, N. A. G., & Berkovsky, S. (2022). Modeling hybrid feature-based phishing websites detection using machine learning techniques. *Annals of Data Science, 9*(3), 705–727. https://doi.org/10.1007/s40745-021-00343-7

5. Benavides-Astudillo, E., Fuertes, W., Sánchez-Gordon, S., Núñez-Agurto, D., & Rodríguez-Galán, G. (2023). A phishing-attack-detection model using natural language processing and deep learning. *Applied Sciences, 13*(9), 5275. https://doi.org/10.3390/app13095275