Project plan

Phishing Link AI Detection

Development

Jonathan Christyadi - 502705

Table of Contents

Introduction

With the rapid expansion of digital technologies, our lives have become more interconnected and convenient. However, this digital revolution also brings along its fair share of risks, particularly in the form of cyber threats. Among these threats, phishing attacks have emerged as a prevalent and deceptive form of cybercrime. Phishing relies on tricking individuals into sending sensitive information, such as passwords or financial details, through deceptive emails or websites.

Phishing is not just a minor inconvenience; it is a significant issue in cybersecurity with substantial economic and personal ramifications. According to recent reports, phishing attacks have caused billions of dollars in losses annually, affecting both individuals and organizations. These attacks exploit human psychology and trust, making them particularly difficult to defend against using traditional methods. The impact includes financial losses, compromised personal information, and erosion of trust in digital communications.

In response to the escalating threat posed by phishing attacks, there is a pressing need for effective cybersecurity solutions. This project proposes the development of an AI-powered phishing link detection system. By leveraging advanced machine learning techniques, the goal is to enhance our ability to detect and mitigate phishing attacks. Ultimately, the aim is to empower users and organizations with proactive defenses against phishing attempts, thereby safeguarding digital security and privacy.

Background

Phishing attacks have evolved considerably over time, exploiting vulnerabilities in human behavior and online security measures. Traditional methods of detecting phishing, such as maintaining lists of known malicious URLs or using static rule-based filters, have proven insufficient in keeping pace with the evolving tactics of cybercriminals. Furthermore, the emergence of targeted phishing attacks, known as spear phishing, has further complicated the cybersecurity landscape.

The magnitude of the issue is underscored by the sheer volume and sophistication of phishing attempts. For instance, from 2017 - 2022, the FBI’s Internet Crime Complaint Center (IC3) received nearly 3.26 million complaints about phishing attacks, with reported

losses exceeding $27.6 billion. These figures highlight the critical need for advanced, adaptable security measures.

Artificial intelligence, particularly machine learning and natural language processing, offers promising avenues for combating phishing attacks. By analyzing patterns in online behavior and content, AI algorithms can detect subtle signs of phishing activity, enabling early detection and response. Integrating deep learning models and advanced feature extraction techniques has further improved the accuracy and effectiveness of phishing detection systems.

Against this backdrop, this project aims to harness the capabilities of AI technologies to develop a robust phishing link detection system. By combining innovative machine learning algorithms with comprehensive feature extraction methods, the objective is to create a proactive defense mechanism capable of identifying and mitigating phishing attacks in real-time. Through collaborative efforts and interdisciplinary research, we strive to contribute to the ongoing evolution of cybersecurity practices, fostering a safer and more resilient digital environment for all.

Stakeholders

- **Parents and Vulnerable Individuals:** Parents and other vulnerable individuals are often targets for phishing attacks due to their limited technical knowledge and experience with digital security. These stakeholders provide valuable insights into the real-world impact of phishing attacks and the effectiveness of the detection system. Their experiences and feedback can guide the development of more user-friendly and accessible solutions. Engaging with this group through surveys, interviews, and usability testing will ensure that the system meets their needs and effectively protects them from phishing threats.

- **Friends**: This group consists of friends who are passionate about software development and artificial intelligence and are active internet users. These friends play a role in developing and refining the phishing detection system. Their enthusiasm and collaborative spirit are essential for building a robust and accurate model. By working together, sharing ideas, and providing feedback, this group can help drive the project's success and innovation.

Project Goals

- **Implement Data Cleaning Procedures**: Ensure the quality and reliability of the dataset by implementing rigorous data cleaning procedures. This includes addressing missing values, removing duplicates, and standardizing data formats to create a robust foundation for analysis and model development.
- **Conduct Comprehensive Data Analysis**: Utilize advanced data analysis techniques to gain insights into the characteristics of phishing links. Employ visualization methods such as plots and graphs to uncover patterns and trends within the dataset, enabling a deeper understanding of the underlying factors contributing to phishing activities.
- **Develop an AI Model for Phishing Link Detection**: Build a sophisticated AI model capable of accurately recognizing phishing URL links. Utilize state-of-the-art machine learning algorithms and feature extraction techniques to train the model on the cleaned dataset, enabling it to effectively differentiate between legitimate and phishing URLs.
- **Enhance Model Accuracy**: Continuously improve the AI model to enhance its accuracy in identifying phishing URL links. Explore approaches for feature engineering, algorithm selection, and model optimization to achieve superior performance and reliability in real-world scenarios.

Scope

The project includes:	The project does not include:
1 Phishing link detection	1 Automatic phishing input
2 Frontend/ website	2 Maintenance plan

Methodology

Data Collection: Gather a diverse dataset comprising both legitimate and phishing URLs from kaggle or mendeley data (Web page phishing detection - Mendeley Data).

Feature Extraction: Extract relevant features from URLs, including domain characteristics, URL structure, presence of HTTPS, and the page content itself.

Model Development: Train machine learning models, leveraging algorithms such as Random Forest, K-nearest neighbors, Support vector machine, AdaBoost.

Model Evaluation: Assess the performance of the trained models using standard evaluation metrics such as accuracy, precision, recall, F1-score, and confusion-matrix.

Domain Understanding

For this project, I will use Field and Library Research from ICT Research Methods, especially community research and exploratory data analysis.

Research questions:

1. What specific URL features are indicative of phishing links?
- How do phishing URLs differ from legitimate URLs in terms of structure and syntax?
- Are there discernible patterns in domain names commonly used in phishing links?

2. How do users assess the trustworthiness of URLs, and how does this impact interaction with phishing links?
- What role do domain names and website content play in shaping users' perceptions of trustworthiness?

3. Do phishing tactics and link characteristics vary across different cultures or regions?
- Are there region-specific trends in phishing link creation and distribution?

4. How do phishing attack techniques evolve over time?
- What are the emerging trends in phishing tactics and strategies?

5. How can feature engineering optimize the performance of machine learning models in detecting phishing links?
- Which features are most informative for accurately identifying phishing URLs?
- How does the selection of features impact the model's precision and recall?

6. How do cultural and societal factors influence users' perceptions of trustworthiness in URLs?

- How do users' past experiences and exposure to phishing attacks affect their trust assessment of URLs?

7. What are the long-term consequences of falling victim to phishing scams on individuals' online behaviors and trust in digital platforms?
- How do victims of phishing attacks experience psychological distress, trust erosion, and feelings of vulnerability?

Answers:

1. Specific URL Features Indicative of Phishing Links

Phishing attacks represent a significant cybersecurity threat, exploiting human vulnerabilities through deceptive URLs. Research by Hossain et al. (2019) highlights specific URL features that are indicative of phishing links. These features include domain age, presence of HTTPS, use of subdomains, URL length, and presence of non-standard characters. Phishing URLs often deviate from standard syntax, exhibiting irregularities such as misspellings, hyphens, and excessive subdomains (Caviglione et al., 2019). In contrast, legitimate URLs adhere to consistent structure and syntax conventions.

2. Discernible Patterns in Domain Names Commonly Used in Phishing Links

Bansal et al. (2018) observe discernible patterns in domain names commonly used in phishing links. These patterns include variations of popular domains, misspelled or spoofed domain names, and domains with high entropy or randomness. Such patterns contribute to the deceptive nature of phishing URLs, aiming to trick users into divulging sensitive information.

3. User Assessment of URL Trustworthiness and its Impact on Interaction with Phishing Links

Users assess the trustworthiness of URLs based on factors such as domain reputation, website content, and presence of security indicators. Kim et al. (2020) suggest that trust assessment significantly influences user interaction with phishing links, impacting their likelihood to click on or engage with suspicious URLs. Domain names and website content play a crucial role in shaping users' perceptions of trustworthiness (Miyazaki & Fernandez, 2001). A legitimate domain with authentic content is more likely to be perceived as trustworthy, while suspicious domain names or low-quality content may raise red flags for users.

4. Variability of Phishing Tactics and Link Characteristics Across Cultures and Regions

Cultural and regional differences can influence phishing tactics and link characteristics due to varying cultural norms, language preferences, and regional cybersecurity awareness. Research by Zhu et al. (2012) suggests that region-specific trends in phishing link creation and distribution may emerge based on local socio-economic factors and internet usage patterns.

5. Evolution of Phishing Attack Techniques Over Time

Phishing attack techniques evolve over time to circumvent security measures and exploit new vulnerabilities. Jakobsson et al. (2018) highlight emerging trends in phishing tactics, including spear phishing, social engineering techniques, and leveraging current events or crises for deception.

6. Importance of Feature Engineering in Optimizing Machine Learning Models for Phishing Link Detection

Feature engineering plays a crucial role in optimizing machine learning models for phishing link detection. Zheng and Casari (2018) emphasize the selection of informative features such as domain age, URL length, and lexical analysis of URL components. The choice of features directly impacts the model's precision and recall, influencing its ability to accurately differentiate between phishing and legitimate URLs.

7. Psychological Consequences of Falling Victim to Phishing Scams

Falling victim to phishing scams can have long-term psychological consequences, including emotional distress, trust erosion in online interactions, and heightened vigilance in engaging with unfamiliar websites or links (Fogg et al., 2018). These consequences underscore the importance of effective cybersecurity measures and user education in mitigating the impact of phishing attacks.

References:

Zheng, X., & Casari, A. (2018). Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists. O'Reilly Media.

Jansen, B. J., Zhang, M., Sobel, K., & Chowdury, A. (2017). Phishing on Twitter: Social media trolling for phishing. Journal of the Association for Information Science and Technology, 68(5), 1119-1130.

Guerra, N., Tan, S. S. L., & Sinha, R. (2019). Understanding phishing from the perspective of cybersecurity experts and internet users: A comparative study. Computers & Security, 84, 148-160.
Kumar, P., & Aksit, M. (2018). Social engineering attacks and phishing: A survey. Computers & Security, 78, 320-337.

Data Sourcing

Objective:

The objective of the data sourcing phase is to gather a comprehensive and diverse dataset that can be used to train and validate our AI model for phishing link detection. This dataset should include both legitimate and phishing URLs, with detailed features that can aid in classification.

Data Requirements:

Phishing URL links with corresponding labels indicating whether they are phishing links or legitimate links. Data encompassing the content of the URLs, including textual information, website structure, and any additional metadata that may aid in classification.

Data Requirements:

- Phishing URLs: Collect phishing URLs with labels indicating they are malicious.
- Legitimate URLs: Collect legitimate URLs with labels indicating they are safe.
- URL Features: Gather data on URL structures, domain characteristics, presence of HTTPS, textual content, and user interaction patterns.

Data Legality and Ethics:

Use only publicly available data or data with appropriate licenses to ensure compliance with legal and ethical standards.
Respect the terms of use stipulated by the data sources, including any restrictions on data usage or redistribution.

Data Diversity:

Gather phishing links from various geographical regions and time periods to ensure the dataset's diversity and representativeness of global phishing trends.

Version Control:

Use GitLab for version control, facilitating collaborative development and tracking changes throughout the project lifecycle.

Iterative Process:

- Implement an iterative approach to model development, continuously refining and improving the model to enhance its accuracy.
- Optimization of feature selections, ensuring the dataset's quality, and fine-tuning hyperparameters are integral components of the iterative process.

AI in Phishing Detection

Various AI technologies are already in use for phishing detection.

- **Google Safe Browsing:**

Uses machine learning to identify and block phishing websites. It analyzes URL patterns, website behaviors, and other factors to detect malicious sites.

- **Microsoft Defender SmartScreen:**

Employs AI to protect users from phishing attacks by analyzing URLs and detecting potentially dangerous websites. It integrates with browsers to provide real-time protection.

- **PhishAI**:

A machine learning-based solution that analyzes email content and URLs to detect phishing attempts. It leverages natural language processing to identify suspicious language and patterns.

Analytic Approach

The analytic approach involves a structured methodology to develop, train, and evaluate the AI model for phishing link detection. The process includes the following steps:

1. Target Variable:

The target variable is the link label, distinguishing between phishing links and legitimate links. This binary classification problem serves as the foundation for model development and evaluation.

2. Data Preparation:

- Data Cleaning: Address missing values, remove duplicates, and standardize data formats to create a robust dataset.
- Feature Extraction: Extract relevant features from URLs, including domain characteristics, URL structure, presence of HTTPS, and the content page itself.

3. Model Development:

Train machine learning models using algorithms such as Random Forest, K-nearest neighbors, Support Vector Machine, and AdaBoost. Experiment with various algorithms to determine which best suits the project's objectives and provides the best result in phishing link detection.

4. Model Evaluation:

Assess the performance of the trained models using standard evaluation metrics such as accuracy, precision, recall, and F1-score. Implement cross-validation to ensure the model's robustness and generalizability.

5. Continuous Improvement:

Continuously improve the AI model by exploring approaches for feature engineering, algorithm selection, and model optimization. Conduct thorough testing and validation to refine the model and enhance its accuracy in identifying phishing URL links.

6. Interpretation and Insights:

Interpret the results to understand how well the models capture phishing link characteristics. Identify insights and patterns revealed by the models, and use these to further refine the approach.

7. Comparison of Algorithms:

Compare the performance of different algorithms and models to identify the most effective solution. Evaluate the advantages, disadvantages, accuracy, and other performance metrics of each algorithm to select the best approach for phishing detection.