University of Barisal

PHISHING WEBSITE DETECTION BY USING MACHINE LEARNING TECHNIQUES

Project Proposal

Submitted to

Dr. Tania Islam

Assistant Professor

Department of Computer Science and Engineering
University of Barishal

Submitted By

Popy Das

Roll: 09-001-05

Department of Computer Science and Engineering
University of Barishal

Contents

Introduction

Phishing attacks are among the most prevalent and harmful cyber threats, exploiting the trust of online users to fraudulently obtain sensitive information such as usernames, passwords, and financial details. These attacks often employ deceptive techniques, including spoofed emails and fake websites that closely mimic legitimate ones. With the increasing reliance on online services like banking, education, entertainment, and social networking, phishing has become a significant concern, impacting individuals, businesses, and even governments.

Traditional methods of phishing detection, such as blacklists and heuristic-based systems, face limitations in identifying newly emerging phishing sites. In contrast, machine learning (ML) techniques provide a more dynamic and efficient approach to phishing detection. By training models on datasets that contain both phishing and legitimate websites, ML algorithms can identify suspicious patterns and features in URLs and website content indicative of phishing attempts.

The objective of this project is to develop an effective phishing detection system using machine learning techniques. A comprehensive dataset of phishing and legitimate websites will be gathered, and relevant features will be extracted. Various ML models and deep neural networks (DNNs) will be trained to predict phishing sites, and their performance will be measured and compared to determine the most effective solution. This project aims to improve the early detection of phishing websites, reducing the risk of users falling victim to such attacks.

Background

Phishing attacks are a growing cyber threat, targeting individuals and organizations to steal sensitive information such as credit card details, login credentials, and personal data. The frequency and sophistication of phishing incidents have been steadily increasing over the years. In 2008, the Anti-Phishing Working Group recorded 51,401 phishing websites, and by 2016, global losses due to phishing reached an estimated $9 billion, according to RSA Security Inc [1, 2]. These statistics clearly demonstrate that existing anti-phishing solutions are inadequate in effectively combating phishing attacks, which continue to evolve rapidly, exploiting users' trust in familiar

websites and services.

The detection of phishing websites is critical, particularly in sectors such as online banking and trading, where users are more vulnerable to these attacks. Many users believe that anti-phishing techniques can protect them, but the limitations of current methods often allow phishing sites to bypass defenses. Traditional approaches, such as URL blacklists, only detect known phishing sites, leaving users exposed to newly created threats. Another approach is meta-heuristics, which collect features from URLs and websites, such as URL length, domain age, and website content, before applying classification techniques to determine if a site is phishing or legitimate [5]. However, these methods are not foolproof and can be circumvented by more sophisticated phishing techniques.

Machine learning (ML) has emerged as a more effective approach to phishing detection. ML models analyze a range of features extracted from websites, learning relationships between them to identify patterns that can distinguish phishing websites from legitimate ones. These models offer greater flexibility and accuracy, as they can be trained on large datasets and continuously updated to recognize new phishing tactics. Machine learning techniques not only enhance the detection of phishing websites but also reduce false positives, providing a more reliable defense against phishing. By incorporating ML-based methods, phishing detection systems can adapt to the evolving nature of phishing attacks, offering improved protection for online users.

Objective

- Identify phishing websites that mimic legitimate URLs and webpages.
- Develop machine learning models and deep neural networks (DNNs).
- Assemble a dataset comprising both phishing and Legitimate URLs.
- Extract relevant features from URLs and website content.
- Evaluate and compare the performance of each model.

Research Questions

- What features distinguish phishing URLs from benign URLs?

- How effective are various machine learning models in detecting phishing websites?
- How can the adaptability of machine learning models be enhanced to recognize new phishing tactics?

Motivation

The increasing prevalence of phishing attacks poses a significant threat to individuals, organizations, and entire economies. As more people rely on online services for banking, shopping, and communication, the potential for cybercriminals to exploit vulnerabilities has escalated dramatically. Phishing attacks not only result in financial losses but also undermine trust in digital platforms, which is essential for the growth of e-commerce and online services.

The motivation behind this project stems from the urgent need for more effective and adaptive phishing detection mechanisms. Traditional methods, such as blacklist-based systems, are insufficient in addressing the evolving tactics of cybercriminals. By leveraging machine learning and deep learning techniques, this project aims to enhance the detection of phishing websites by identifying subtle patterns and features that distinguish them from legitimate sites. Improving the accuracy and responsiveness of phishing detection systems is crucial for protecting users' sensitive information and maintaining trust in digital interactions.

Related Study

The rapid advancement of malicious software presents significant challenges within the cybersecurity landscape, necessitating more effective detection mechanisms. Traditional methods for malware detection, such as graph-based, rule-based, and entropy-based approaches, have shown to be impractical when addressing the complexities and dynamics of new malware variants. As cyber threats continue to evolve, these conventional techniques struggle to keep pace, underscoring the necessity for innovative solutions. Machine learning (ML) techniques have emerged as a promising alternative, offering enhanced capabilities for detecting and mitigating emerging malware threats.

Beyond the realm of supervised ML approaches, researchers have also explored unsupervised techniques to bolster detection efforts. Among the

widely employed classification methods are Naïve Bayes (NB), Support Vector Machines (SVMs), Random Forests (RFs), and AdaBoost [6]. Each of these methods presents unique advantages and limitations, making it imperative to evaluate their effectiveness in various contexts. Moreover, ensemble learning techniques, which combine multiple models to improve overall performance, have been shown to achieve superior results in combating malware attacks. This highlights the critical need for the development of sophisticated detection methods capable of adapting to the changing landscape of cyber threats. To facilitate effective defenses against website attacks, automated and cognitive-based analysis systems are proposed. These systems would benefit from a continuous influx of updated information regarding malware behaviors, patterns, and variants, allowing for more timely and accurate detection.

In the specific context of phishing detection, most existing machine learning techniques rely heavily on feature extraction, which has proven to yield high accuracy rates. Research conducted by Zhu et al. indicates that over 200 distinct features can be extracted from web data, providing a rich foundation for analysis [8]. However, an excessive number of features can complicate the design of classifiers and lead to significant overfitting issues, where the model becomes overly tailored to the training data and fails to generalize effectively to new data. This reality highlights the critical importance of optimal feature selection, a significant challenge within traditional machine learning techniques in phishing detection. The present research aims to identify the most relevant features for effectively detecting phishing attempts.

To address this challenge, a proposed methodology known as the "Decision Tree and Optimal Features based Artificial Neural Network" (DTOF-ANN) has been developed [8]. This approach leverages artificial neural networks (ANNs) to construct a robust classifier. Before selecting the optimal features for the model, the importance of each feature is thoroughly evaluated. This evaluation process leads to the formation of an optimal feature vector, which enhances the classifier's performance by focusing on the most relevant data points.

In parallel efforts, Waleed Ali has proposed a systematic procedure for detecting phishing websites utilizing a variety of supervised machine learning techniques. This includes methods such as Radial Basis Function Network (RBFN), Naïve Bayes Classifier (NB), Back-Propagation Neural Network (BPNN), Decision Tree, k-Nearest Neighbors (kNN), Random Forest (RF), and Support Vector Machine (SVM). Ali's approach employs wrapper feature

selection based on these classifiers to optimize detection accuracy. However, research findings indicate that while neural network models can effectively classify phishing attempts, they are prone to underfitting if poorly structured, leading to inadequate model performance [3]. Conversely, if a model is designed to fit every individual data point within the training set too closely, it may suffer from overfitting, rendering it ineffective in real-world applications [4, 7].

The primary objective of this project is to develop an effective phishing detection system that leverages advanced machine learning techniques. To achieve this goal, a comprehensive dataset comprising both phishing and legitimate websites will be meticulously gathered. Relevant features will be extracted from this dataset to inform the training of various machine learning models and deep neural networks (DNNs) aimed at predicting phishing sites. The performance of these models will be rigorously measured and compared to determine the most effective solution for phishing detection. Ultimately, this project seeks to enhance early detection capabilities, thereby significantly reducing the risk of users falling victim to phishing attacks.

Research Methodoloy

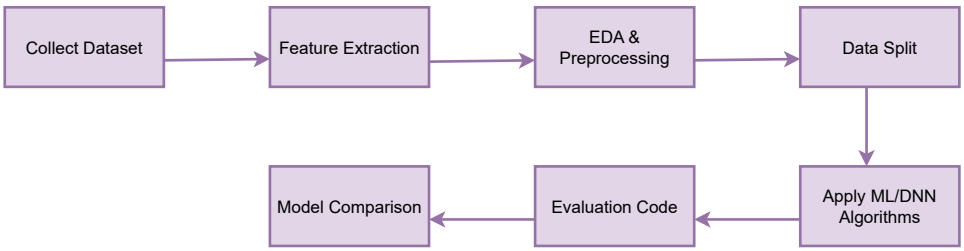

Figure 1: Step by Step Research Methodology

The methodology for detecting phishing websites involves several systematic steps to ensure effective data collection, analysis, and model evaluation. First, a dataset comprising both phishing and legitimate websites is collected from open-source platforms. This diverse dataset is essential for training machine learning models to recognize patterns associated with phishing attacks. Next, code is developed to extract relevant features from the URL database, focusing on critical characteristics such as domain names, URL lengths, and

the presence of special characters, which are indicative of phishing behavior. Following feature extraction, exploratory data analysis (EDA) techniques are employed to analyze and preprocess the dataset. This step includes visualizing data distributions, identifying missing values, and normalizing features to enhance the dataset's quality. After preprocessing, the dataset is divided into training and testing sets, typically using an 80/20 split, to evaluate model performance accurately. Selected machine learning algorithms, such as Support Vector Machine (SVM) and Random Forest, along with deep neural network models like Autoencoders, are then applied to the training data. Each model is trained to identify phishing URLs based on the extracted features. To assess the effectiveness of the models, code is written to evaluate their performance using various accuracy metrics, including precision, recall, and F1-score. Finally, the results from the trained models are compared, highlighting which algorithm performs best in detecting phishing attempts based on accuracy metrics. This comprehensive methodology ensures a robust approach to phishing detection, combining data-driven insights with machine learning techniques.

Time Schedule

project Proposal	10 Days
Data Collection	1 Week
project Development	2 Weeks
Testing and Final documentation	1 week

Expected Outcome

The expected outcome of this methodology is a robust machine learning model capable of accurately identifying phishing websites among legitimate ones. By employing a diverse dataset and extracting relevant features, the model aims to achieve high precision and recall rates, minimizing false positives and negatives. Additionally, the comparative analysis of different algorithms, such as SVM, Random Forest, and Autoencoders, is anticipated to reveal the most effective approach for phishing detection. Ultimately, the successful implementation of this methodology is expected to enhance cybersecurity measures, providing users and organizations with a reliable tool to

protect against phishing attacks and contribute valuable insights to the field of online security.

	ML Model	Train Accuracy	Test Accuracy
3	XGBoost	0.868	0.857
2	Multilayer Perceptrons	0.866	0.854
4	AutoEncoder	0.810	0.810
1	Random Forest	0.820	0.809
0	Decision Tree	0.816	0.803
5	SVM	0.806	0.786

Figure 2: Expected outcome

Conclusion

This project proposal outlines a comprehensive approach to detecting phishing websites using advanced machine learning and deep learning techniques. By leveraging a robust dataset of both phishing and legitimate URLs, the project aims to develop models that effectively distinguish between the two, thereby enhancing online security. The proposed methodology includes critical steps such as data collection, feature extraction, exploratory data analysis, model training, and evaluation, ensuring a systematic and data-driven approach to problem-solving. The anticipated outcomes promise not only to yield a high-performing detection system but also to contribute valuable insights into phishing behavior and patterns. Ultimately, this project seeks to equip users and organizations with the necessary tools to mitigate phishing threats, fostering a safer online environment. Through rigorous analysis and model comparison, we hope to identify the most effective strategies for combating phishing, making a significant impact in the realm of cybersecurity.

References

[1] H Bleau. Global fraud and cybercrime forecast, 2016, 2017.

[2] Kang Leng Chiew, Choon Lin Tan, KokSheik Wong, Kelvin SC Yong, and Wei King Tiong. A new hybrid ensemble feature selection framework for machine learning-based phishing detection system. *Information Sciences*, 484:153–166, 2019.

[3] Stefan Duffner and Christophe Garcia. An online backpropagation algorithm with validation error-based adaptive learning rate. In *International Conference on Artificial Neural Networks*, pages 249–258. Springer, 2007.

[4] Rami M Mohammad, Fadi Thabtah, and Lee McCluskey. Predicting phishing websites based on self-structuring neural network. *Neural Computing and Applications*, 25:443–458, 2014.

[5] Ozgur Koray Sahingoz, Ebubekir Buber, Onder Demir, and Banu Diri. Machine learning based phishing detection from urls. *Expert Systems with Applications*, 117:345–357, 2019.

[6] Jagsir Singh and Jaswinder Singh. A survey on machine learning-based malware detection in executable files. *Journal of Systems Architecture*, 112:101861, 2021.

[7] Fadi Thabtah, Rami M Mohammad, and Lee McCluskey. A dynamic self-structuring neural network model to combat phishing. In *2016 international joint conference on neural networks (ijcnn)*, pages 4221–4226. IEEE, 2016.

[8] Erzhou Zhu, Yinyin Ju, Zhile Chen, Feng Liu, and Xianyong Fang. Dtof-ann: an artificial neural network phishing detection model based on decision tree and optimal features. *Applied Soft Computing*, 95:106505, 2020.