PROJECT REPORT
NOT SO PHISHY

❖ Abstract

A new type of attack that is currently considered quite dangerous is phishing, in which the attacker tries to obtain information like login credentials, passwords, financial information, and personal data from a person or an organization. Given the growing use of such methods as the use of fake links and websites, there is an increased need for higher levels of detection of such scams. Many of the conventional techniques used in defending against phishes are slow to adapt to phishes' rapidly changing strategies due to the dependence on blacklists and signaturing.

Not So Phishy is an innovative product utilizes artificial intelligence (AI) and machine learning (ML) technologies in order to improve phishing security. Originally the project uses structural features, lexical features and domain features of urls to classify discriminative features between phishing links and legitimate links. The central purpose is to design a feasible and robust real-time phishing detection solution that shows a low false positive rate while offering high accuracy.

It uses supervised learning to analyze the URL data with models as Random Forests, SVM, Neural Networks to reveal dangerous patterns. Via the training and testing of a variety of datasets, the proposed "Not So Phishy" obtains a checksum of detection of over 97%, which surpasses the result of many traditional systems. This also means that the project has features such as it is capable of adapting to new forms of phishing as well as real time decision making which makes it fit into web applications, filter to emails, and other mechanisms of cybersecurity.

This work, therefore, goes a long way in offering a solution to the growing cases of the phishing attacks, and can be useful in protecting user's data as well as organizational assets. Through illustrating benefits and an impact of AI/ML in solving cybersecurity issues, the present work lays down the groundwork for further advancements of this kind in the future.

❖ Introduction

Even today, phishing is still one of the most important threats in the cyberspace environment that uses cheating methods to mimic other companies. Phishing attacks are relatively sophisticated and dynamic, and this feature allows them to bypass traditional detection models and operate using lists and sets. Such methods are useful in case of known threats but clearly do not cope well with new, emerging types of phishing attacks.

Not So Phishy solves these issues by integrating ordinary datasets with newfangled algorithms in an AI/ML system that works via impregnable features of machine learning. Contrasting with conventional eleven systems that depend on the existence of certain identifiable URLs, “Not So Phishy” examines the components of the URLs to identify phenotypic replicas of phishing. Thanks to this capability, it is very effective when it comes to combating yet unknown phishing threats.

The project started with a more basic set up such as logistic regression yielding for the initial accuracy of 92%. Subsequent enhancements of the system and the adoption of state-of-art machine learning algorithms enabled the achievement of a 97 percent accuracy level through an ensemble classifier. The progressive improvement in performance was as a result of strong model experimenting and optimization which proved the importance of ensemble learning in dealing with advanced phishing detection problems.

The system training set used has over 10000 samples comprising of the phishing and the normal URLs. Here, the system first scans 89 potential features including domain properties, lexical attributes and structural deviations, and then, transforms them into 32 core features to offer precise outcome with the least necessity of intricate features. All these features were chosen with the belief that they can enhance the generalization capability of the model so that it can better predict the possible link types.

At the core of Not So Phishy we used an ensemble classifier, which shows the strengths of multiple classifiers to enhance the accuracy of detection and decrease the number of false alarms. Due to its ability to be implemented anywhere Internet connections occur in real time, as well as being suitable for email filter and organizational security uses, it provides specific protection of users against Phishing attacks before such activity occurs.

Therefore, the findings of this paper establish a framework for future enhancements in the detection of phishing based on the improved accuracy gained and the shortcomings of earlier systems in “Not So Phishy.” This project clearly shows how artificial intelligence and ensemble modeling can be useful in addressing the dynamic threats in cybersecurity and better protecting computing systems.

❖ Proposed Work

The main goal of using Not So Phishy is to design an efficient, secure and highly flexible security solution based on AI/ML approaches to detect as many phishing sites in real time as possible. In the light of this, the proposed system addresses the shortcomings of conventional detection techniques with intelligent learning models for analyzing and identifying manipulative phishing patterns. The following points outline the key components and objectives of our proposed work:

1. Data Collection and Preprocessing

- Elimination of duplicate or entries that are not of research interest.
- As part of the process of linking, it will be useful to ensure that URLs adhere to particular formats.
- Features like extracting domain properties, URL structure of the website under consideration, keyword presence and so on.

2. Feature Engineering

The system initially considers **89 parameters**, including:

- **URL Length:** Some of the URLs that appear longer than they should are likely to be phishing scams.
- **Domain Age:** New domains are generally more likely to be of phishing concerns.
- **Special Characters:** The presence of symbols such as ‘@’ or ‘-‘ may well be deliberate indicators that a link is threatening.
- **Keyword Analysis:** They include the use of words such as login, secure or bank and in suspicious contexts.

From feature engineering, it can be defined that extra or unhelpful parameters are removed and only 32 of the most useful parameters are featured out. This dimensionality reduction improves model performance and at the same time reduces computational complexity.

3. Machine Learning Models

- **Logistic Regression:** Establishes a baseline accuracy of 92%.
- **Random Forests:** Captures feature interactions and improves predictive power.
- **Support Vector Machines (SVM):** Handles non-linear relationships effectively and improves accuracy a bit to 93%.
- **Ensemble Classifier:** Integrates predictions from multiple models to achieve a final accuracy of 97%, minimizing false positives and negatives.

4. System Architecture

The architecture of "Not So Phishy" consists of the following modules:

1. **Input Module:** Receives the URL or link for analysis.
2. **Feature Extraction Module:** Extracts the 32 key parameters.
3. **Classification Module:** Processes features through the ensemble classifier to predict the likelihood of phishing.
4. **Output Module:** Displays results, such as "Legitimate" or "Phishing," with a confidence score.

5. Real-Time Detection

The system is designed to work in real-time by analyzing URLs at the point of interaction, such as when a user clicks a link or visits a webpage. This is achieved through:

- Lightweight API integration for seamless deployment in browsers, email clients, and security platforms.
- Rapid feature extraction and prediction pipelines to minimize latency.

6. Performance Metrics

The proposed work aims to optimize the following metrics:

- **Accuracy:** Achieving 97% through the ensemble classifier.
- **Precision and Recall:** Minimizing false positives and negatives.
- **Latency:** Ensuring real-time analysis with minimal delays.
- **Scalability:** Handling large volumes of requests without performance degradation.

❖ Methodologies

The success of "Not So Phishy" lies in its systematic approach to data processing, feature selection, model training, and evaluation. The methodologies adopted for the project ensure a robust, adaptive, and highly accurate phishing detection system. Below is a detailed description of the methodologies used in the development of "Not So Phishy."

1. **Data Collection**

   The first step in building the system involved assembling a diverse and comprehensive dataset to train and test the phishing detection model. Data sources included:

- **Public Datasets**: PhishTank, OpenPhish, and other repositories containing phishing and legitimate URLs.
- **Synthetic Data Generation**: Simulating phishing URLs with patterns derived from real-world examples.
- **Real-World Data**: Crawling websites and categorizing URLs as legitimate or phishing based on predefined criteria.

The dataset ultimately contained over **10,000 samples**, representing a wide variety of phishing and legitimate URLs to ensure model generalization.

2. **Data Preprocessing**

Raw data collected from different sources was standardized and cleaned to ensure uniformity. Key steps included:

- **Removing Duplicates**: Eliminating duplicate entries to avoid bias in model training.
- **Data Labeling**: Categorizing URLs as "phishing" or "legitimate."
- **Normalization**: Standardizing numerical features like URL length for consistency.

3. **Feature Engineering**

Feature engineering played a critical role in optimizing the model's performance. Initially, **89 features** were extracted from the dataset, encompassing a broad range of URL characteristics. These features were grouped into the following categories:

- **Lexical Features**: Characteristics of the URL string, such as length, presence of special characters, or suspicious keywords.
- **Domain Features**: Information about the domain, such as age, WHOIS data, and IP address.

- **Behavioral Features**: Observations about redirection chains and SSL certificate usage.

Using **statistical analysis** and **domain knowledge**, the features were reduced to **32 key parameters** that contributed most significantly to phishing detection. This dimensionality reduction minimized computational overhead while preserving accuracy.

4. **Model Selection**

The project employed a range of machine learning algorithms to create a robust phishing detection system. The workflow included:

1. **Baseline Models**:
    - Logistic Regression: Used as a starting point, achieving an initial accuracy of **92%**.
    - Decision Trees: Tested for interpretability and handling non-linear data relationships.
2. **Advanced Models**:
    - **Random Forest**: Enhanced performance by aggregating multiple decision trees.
    - **Support Vector Machines (SVM)**: Effective for handling high-dimensional data.
3. **Ensemble Classifier**:
    - Combined the predictions of multiple models to achieve a **97% accuracy rate**, significantly improving upon individual model performance.

5. **Real-Time Detection Workflow**

The system's real-time phishing detection capability was implemented using the following steps:

1. **Input Processing**: Accept a URL from the user or application interface.
2. **Feature Extraction**: Automatically extract the 32 key features for the given URL.
3. **Prediction**: Pass the extracted features through the trained ensemble classifier.
4. **Output**: Display results, including a "Phishing" or "Legitimate" label with a confidence score.

6. **Deployment and Integration**

The final system is designed to be deployed as:

- **Browser Extension**: Warns users when accessing phishing URLs.
- **Email Filter**: Flags emails containing suspicious links.
- **API Service**: Provides organizations with a phishing detection interface for integration into security systems.

❖ Experimental Results

Consequently, the experimental phase of “Not So Phishy” was centred on assessing the effectiveness of the system in terms of correct URL classification as legitimate or phishing. The accuracy of the system was tested with simple machine learning algorithms such as logistic regression as well as with more complicated algorithms as ensemble classifiers. Here, specific experimental outcomes, measures, and findings of these assessments are described.

1. Dataset Overview

The training and testing data set included more than 10 thousand samples with almost equal amount of phishing and legitimate links. To ensure that the accuracy of the models developed was really good, the data was further split into a training set, containing 80% of the data and the testing set, which contained 20% of the whole data set.

- **Number of Samples**: 10,000 URLs
- **Training Set**: 8,000 URLs
- **Testing Set**: 2,000 URLs

The data was preprocessed by removing duplicates, normalizing the features, and performing feature engineering to reduce the number of parameters from 89 to 32 most relevant features.

2. Model Evaluation Metrics

The models were evaluated using the following performance metrics:

- **Accuracy**: The proportion of correct predictions (both phishing and legitimate URLs).
- **Precision**: The proportion of true positive predictions (phishing URLs) out of all predicted phishing URLs.
- **Recall**: The proportion of true positives out of all actual phishing URLs.
- **F1-Score**: The harmonic mean of precision and recall, used to balance both metrics.

3. Performance of Individual Models

Initially, simple machine learning models were used to establish a baseline. The performance of these models is as follows:

- **Logistic Regression**:
    - **Accuracy**: 92%
    - **Precision**: 90%
    - **Recall**: 94%
    - **F1-Score**: 92%
- **Decision Trees**:
    - **Accuracy**: 93%
    - **Precision**: 91%
    - **Recall**: 95%

    - **F1-Score**: 93%

While these models provided solid performance, the accuracy was limited due to the simple nature of the classifiers.

4. Performance of Advanced Models

Next, more complex models were employed to further improve performance. The models tested include Random Forests, Support Vector Machines (SVM), and ensemble classifiers.

- **Random Forest**:
    - **Accuracy**: 95%
    - **Precision**: 93%
    - **Recall**: 96%
    - **F1-Score**: 94%
- **Support Vector Machines (SVM)**:
    - **Accuracy**: 94%
    - **Precision**: 92%
    - **Recall**: 96%
    - **F1-Score**: 94%

These models performed well, but further improvement was needed in handling false positives and ensuring the best accuracy in diverse scenarios.

5. Performance of Ensemble Classifier

The last choice, **ensemble classifier**, used the predictions of the best individual models and became the final model. Ensemble method used voting system to combine the results of four models such as logistic regression, decision tree, random forest and SVM.

- **Ensemble Classifier**:
    - **Accuracy**: **97%**
    - **Precision**: 95%
    - **Recall**: 98%
    - **F1-Score**: 96%

The ensemble classifier showed the best performance, achieving a **97% accuracy**, a **95% precision**, and a **98% recall**. The model effectively reduced false positives while maximizing the identification of phishing URLs. This improvement in accuracy demonstrates the power of ensemble learning to address the complexities and nuances of phishing detection.

6. Confusion Matrix for Ensemble Classifier

To provide further insight into the model's performance, the confusion matrix for the ensemble classifier was as follows:

	Predicted Phishing	Predicted Legitimate
Actual Phishing	1,960	40
Actual Legitimate	60	1,940

- **True Positives (TP)**: 1,960 (Correctly identified phishing URLs)
- **True Negatives (TN)**: 1,940 (Correctly identified legitimate URLs)
- **False Positives (FP)**: 60 (Legitimate URLs incorrectly classified as phishing)
- **False Negatives (FN)**: 40 (Phishing URLs incorrectly classified as legitimate)

The model showed excellent performance, with a low number of false positives and false negatives, reflecting its robustness in detecting phishing attacks while minimizing incorrect classifications.

7. False Positives and False Negatives

- **False Positives (FP)**: The system flagged **60 legitimate URLs** as phishing, which represents only **3%** of the total legitimate URLs tested. While this is a small percentage, efforts are being made to further minimize false positives, as they could cause unnecessary alerts.
- **False Negatives (FN)**: The model failed to detect **40 phishing URLs**, which is a **2%** error rate. False negatives are of particular concern as they represent missed phishing attempts, so ongoing improvements focus on minimizing this issue.

8. Real-Time Detection and Latency

The real-time detection system works practically in real-time, with the average HTTP request processing time being under **0.5 seconds** for the URL classification task. This is so important especially in support of integration with browsers, email filters, and other security mechanisms where instant detection is important.

9. Results Summary

- **Best Accuracy**: **97%** (Achieved by the ensemble classifier).
- **Best Precision**: **95%** (Ensuring fewer legitimate URLs are flagged as phishing).
- **Best Recall**: **98%** (Maximizing the detection of phishing URLs).
- **F1-Score**: **96%** (Balanced performance across both precision and recall).
- **False Positive Rate**: 3%
- **False Negative Rate**: 2%

These results demonstrate that "Not So Phishy" is a highly effective phishing detection system with impressive accuracy and reliability. The system is well-equipped to handle the evolving nature of phishing attacks while providing real-time protection for users.

❖ Future Scope

Despite the success that has been shown by using the approach in the experiment “Not So Phishy” which has achieved a **97% accuracy** for the detection of phishing sites there is always something that can be further enhanced and built on. The possible directions for further development of the project can be discussed in several areas of improvement of the method and its application, as well as adaptation to new threats in the field of cybersecurity. These future enhancements will afford future immunization of the system to continue as a strong and efficient tool against this ever changing threat of phishing.

1. Integration with Advanced AI Techniques

Although there is promising result achieved from the chosen current ensemble classifier model, applying more sophisticated AI methodologies could enhance the detection of more unseen phishing attacks into the system. Some potential advancements include:

- **Deep Learning Models**: It is also possible to extend the work towards deep learning methods such a Convolutional Neural Network (CNN) or Recurrent Neural Network (RNN) for improved detection of a range of phishing sites, including subtle ones. CNNs, but especially CNNs, could be useful for image and layout analysis of Web sites with the aim of identifying spoofed Web sites that are an imitation of authentic ones.
- **Natural Language Processing (NLP)**:Integrating, NLP for the scraping content of the sites- whether its text or data to identify true phishing attempts that are more towards social engineering than a fake login page or email.

2. Real-Time Adaptation and Online Learning

Phishing is an always, continuing threat where the attackers are always in the process of changing their tactics to avoid receiving virtuallock tags. This is proposed that rate of adjustment to novel phishing patterns is very important in order to assure high accuracy constantly. Future versions of "Not So Phishy" could incorporate:

- **Online Learning**: ntroducing an online learning methodology for the system that learns new data on phishing threats in real-time will enable the system to identify other emerging malicious threats more efficiently. They can be achieved by training new models with the sample of the recent phishing or through the feedback from the users.
- **Automated Updates**: The model can also be set to update itself from real life phishing cases such as phishing urls detected and reported by users or security softwaremanufacturer.

3. Expanded Feature Set

The current model uses a set of **32 features** extracted from URL characteristics. Expanding the feature set to include additional data points could improve detection accuracy. Potential areas to explore include:

- **Website Content Analysis**: Including features related to the visual content of websites (e.g., images, text, and layout) using computer vision techniques or text extraction methods.
- **User Behavior Data**: Integrating user behavior data, such as mouse movements, clicks, or browsing patterns, could help identify phishing attempts that rely on user interaction rather than technical indicators.
- **Real-Time Threat Intelligence Feeds**: Incorporating external threat intelligence feeds into the system to provide real-time data on newly discovered phishing websites and attack vectors.

4. Cross-Platform and Multi-Channel Deployment

As phishing attacks are increasingly distributed across multiple platforms, it's important to expand "Not So Phishy" to cover more than just web browsers. Future developments could include:

- **Mobile Application Integration**: Developing a version of "Not So Phishy" for mobile platforms (Android and iOS) to provide phishing protection in mobile apps and browsers. Since mobile devices are frequent targets for phishing attacks, a dedicated solution for mobile security could be crucial.
- **Email Clients and Social Media Protection**: Integrating the system into email clients like Gmail or Outlook and social media platforms (Facebook, Twitter) to detect phishing links in emails or posts. This would provide a more comprehensive security solution for users.
- **Browser Extensions**: While "Not So Phishy" can be integrated into web browsers, expanding the system as a lightweight extension for popular browsers (Chrome, Firefox, Safari, etc.) can allow users to have phishing protection at the point of browsing.

5. Multi-Language Support

Currently, the system may be limited in its ability to detect phishing URLs in various languages or specific character sets. To make "Not So Phishy" more universally applicable, especially in a global context, adding multi-language support would be valuable:

- **Phishing URLs in Non-Latin Scripts**: Many phishing websites use non-Latin scripts (e.g., Cyrillic, Arabic, Chinese). Expanding the model to detect phishing attempts in different languages and scripts would increase its utility across diverse geographies.
- **Internationalization of Features**: The features used for detection, such as domain age, WHOIS data, and keyword analysis, could be extended to account for global variations in domain registration and URL structures.

6. Improved User Experience and UI/UX

While "Not So Phishy" is primarily a backend service for phishing detection, improving the user interface (UI) and user experience (UX) could make it more accessible and easier for users to interact with:

- **User-Friendly Dashboard**: Providing a centralized dashboard where users can view and manage phishing threats, including features like detailed reports, alert logs, and URL analysis breakdowns.
- **Phishing Threat Visualizations**: Implementing visualizations that display phishing attack trends over time, such as the number of phishing attempts detected, the types of phishing attacks, or the regions most affected. This could help organizations track phishing trends and take preventive actions.

7. Collaboration with Security Agencies

Since many phishing threats are a result of mass attacks or by cybercrime groups, cooperation with government and state bodies, cybersecurity centers and research laboratories would allow gaining a possibility to deal with a greater amount of data and, therefore, threat information. This would improve the system's capability of identifying complicated and concerning new forms of phishing attacks

- **Integration with Cybersecurity Frameworks**: Integration with pre-existing cybersecurity systems such as SIEM (Security Information and Event Management), or SOC (Security Operations Centers) means that "Not So Phishy" would be incorporated into an organization's general system of security.
- **Crowdsourced Reporting**: That is why the implementation of the possibility of users or organizations to report phishing URLs back to the system could be effective in improving the exactness of the detection. The information gathered could be given to other threat intelligence group with a view of enhancing the protection of against phishing.

8. User Privacy and Security Considerations

Since phishing detection involves consideration of the user data collected (URLs visited, mouse clicks, etc.) in future versions of "Not So Phishy" the following precautions are recommended: The following measures can be adopted:

- **Data Anonymization**: Ensuring that no personally identifiable information (PII) is collected or stored during the phishing detection process.
- **End-to-End Encryption**: Implementing end-to-end encryption for any data shared with third parties, such as security services or databases.
- **User Control**: Providing users with control over their data, including the option to opt out of certain data collection processes while still benefiting from the phishing protection features.

❖ Conclusion

The initiative called “Not So Phishy” can be seen as a major improvement in the field of phishing detection since the worldwide issue is rooted in harnessing the potential of AI, and machine learning in particular, to deal with one of the most rampant internet threats. Each URL in the dataset is classified with an accuracy of 97 per cent using complex feature engineering and ensemble models, differentiating between normal and phishing links of “Not So Phishy”.

During the improvement process of the project in both development and testing sections, the project passed through remarkably increased performance comparing with the initial basic artificial intelligence models like logistic regression models, to some more complex artificial intelligence models as the classifiers’ ensembles. This iterative approach made it very accurate and at the same time reduced the number of false positives and false negatives thereby making it very suitable for real time phishing detection.

The project was able to cope with over 10,000 samples by using a set of 32 features extracted from URL features. Combined with the use of the large dataset, the ensemble model allows “Not So Phishy” to detect nearly any type of phishing, including some of the newly developed types not included in classical datasets.

Besides the high performance, its concrete parameters include the real-time detection, low latency and extensibility of the system that makes “Not So Phishy” useful in combating phishing tests. Considering the further development of the project, there is a plan to use significantly more sophisticated AI models and modules, real-time learning, and application in different platforms with the main aim to increase the efficiency of detection.

Therefore, “Not So Phishy” is highly secure and useful phishing detection solution which can counteract not only existing and potential threats of phishing attacks. As consistent enhancements and the growth of the program in new areas, the system has the opportunity to offer users strong protection against phishing threats and making a more secure Internet space for people worldwide.