PHISHNET: Intelligence System for Phishing Attacks Using Machine Learning

Mr Prasanna K S *1, Dhanyashri B G *2,Nagashree patgar *3,Varshini N *4,Niharika G J *5

*1.Guide ,Department of Computer Science & Engineering ,Malnad College of Engineering,Hassan,Karnataka,India

*2,3,4,5 Student,Department Of Computer Science & Engineering ,Malnad College of Engineering,Hassan,Karnataka,India

***Abstract*-- Phishing attacks remain a significant threat to cybersecurity, targeting individuals and organizations alike. This paper introduces PHISHNET, a comprehensive threat intelligence system that leverages machine learning techniques to detect and mitigate phishing attacks across two communication vectors: URLs, emails. The system employs Gradient Boosting for URL detection, achieving an accuracy of approximately 98%, while Random Forest algorithms are utilized for email, attaining accuracies of approximately 97% and 96%, respectively. By analyzing various features related to URLs and extracting content from emails, PHISHNET provides users with actionable insights to enhance their security posture against phishing threats.**

Keywords - Feature Extraction, URL Analysis, Real-Time Detection, Cybersecurity, Building Classification, Machine Learning

I. INTRODUCTION

The Phishing attacks have emerged as one of the most prevalent threats in the digital landscape, targeting individuals and organizations alike. These malicious attempts often involve deceptive tactics designed to trick users into revealing sensitive information, such as passwords, credit card numbers, and personal identification details. As cybercriminals continuously evolve their strategies, traditional security measures frequently fall short, necessitating the development of more sophisticated detection systems.

In response to this growing challenge, machine learning techniques have gained traction as effective tools for enhancing phishing detection capabilities. By analyzing patterns in data, these algorithms can identify potential phishing attempts with greater accuracy than conventional methods. This paper presents PHISHNET, a comprehensive threat intelligence system that employs machine learning to detect phishing attacks across various communication channels, including URLs, emails.

In email phishing for example, the hacker sends a link via email to the user. When the user clicks on the link and enters their details, the hacker gains access to the entire user's information. According to Google's Transparency Report 2020 and (APWG), new phishing websites were detected every week. Phishing attacks have progressed and enhanced their techniques as the technology progress and research does not stop at developing or improving its techniques in cybersecurity to detect and classify phishing attacks, by developing several techniques. Many researches have demonstrated that approaches used artificial intelligence (AI) specifically,machine learning with the goal to strengthen system security and prevent any network intrusion .

In recent years, most studies has concentrated on applying machine learning and deep learning methods in cybersecurity, as well as, developing new features selection techniques. As a result, several machine learning techniques have proven to be effective in accurately identifying phishing attacks such as: Decision Tree, Random Forest, SVM, Naïve Bayes, Neural Network and so on, PHISHNET aims to enhance user trust and engagement. Additionally, addressing the cold start problem—where new users lack sufficient data for accurate predictions is crucial for maintaining system efficacy.

Moreover, then temporal aspect of phishing detecti on plays a vital role in ensuring relevance and accuracy. As new phishing tactics emerge and old ones become obsolete, it is essential for detection systems to adapt accordingly. By implementing mechanisms that account for the age of data and user interactions, PHISHNET strives to provide timely recommendations that reflect current threats. In summary, this paper outlines the development phishing attacks using advanced machine learning techniques

SYSTEM DESIGN

1. Email Spam Detection

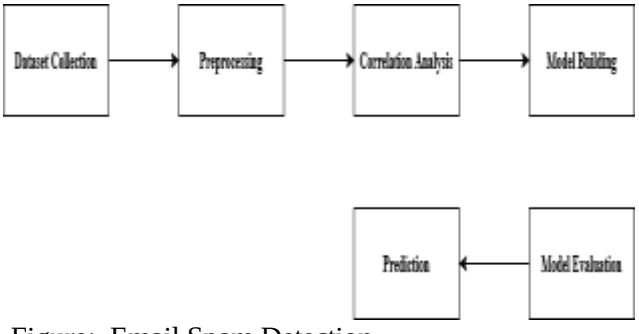

Figure: Email Spam Detection

Email spam detection is a critical area of research aimed at identifying and filtering unsolicited or malicious emails, which can pose threats such as phishing, malware distribution, and data theft. Traditional approaches include rule-based filtering, blacklisting, and heuristic

methods, but these often struggle with evolving spam tactics. Modern techniques leverage machine learning algorithms like Naïve Bayes ,random forest, Support Vector Machines, and deep learning models such as CNNs, RNNs, and transformers, which analyze email content, metadata, and patterns to classify spam with greater accuracy.

1. Support Vector Machine (SVM) -Support Vector Machine (SVM) is a supervised learning algorithm used for classification tasks. It works by finding the optimal hyperplane that best separates the classes in the feature space. In text classification, such as spam detection, SVM is highly effective due to its ability to handle high-dimensional data like TF-IDF vectors. SVM is robust against overfitting, especially when the number of features exceeds the number of samples. It's known for delivering high accuracy and precision in text classification tasks, making it a popular choice for spam filters.

2. Random Forest Classifier- Random Forest is an ensemble learning method that builds multiple decision trees and merges their outputs to improve prediction accuracy and reduce overfitting. In spam detection, each tree in the forest is trained on a random subset of the data and features. This diversity among trees ensures that the model is not overly biased by a particular pattern in the data. When classifying a message, each tree gives a prediction (spam or ham), and the forest selects the class with the majority vote. Random Forest handles large feature spaces and complex interactions well, which is especially useful in text classification using TF-IDF.

3. Multinomial Naive Bayes - Multinomial Naive Bayes is a probabilistic classifier based on Bayes' Theorem and is particularly suited for discrete data such as word counts or TF-IDF values. It assumes that the presence of a particular feature (word) in a class is independent of the presence of others — the "naive" assumption. Despite this simplification, it works surprisingly well for text classification. In spam detection, it calculates the probability of a message being spam or ham based on the frequency of words. For example, if certain words like "free", "offer", or "win" occur more frequently in spam, the model will learn this pattern.

2. Phishing URL Detection

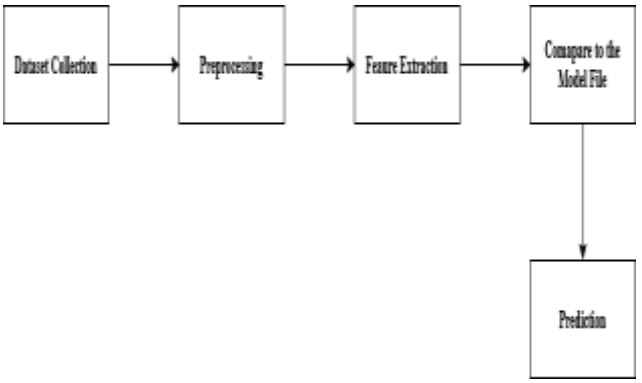

Figure: Phishing URL Detection

Phishing URL detection is a crucial aspect of cybersecurity aimed at identifying malicious links that mimic legitimate websites to steal sensitive user information. Traditional detection methods rely on blacklists and heuristic rules, which are limited in handling novel or obfuscated threats. Machine learning and deep learning approaches have shown significant promise by analyzing URL features such as length, domain patterns, and lexical components, or even learning directly from raw URLs without manual feature extraction. These methods, trained on datasets like PhishTank and OpenPhish, demonstrate improved accuracy and adaptability.

However, challenges such as adversarial evasion techniques, dataset imbalance, and the need for real- time detection remain. Future advancements may involve hybrid models, continual learning, and integration with broader threat intelligence systems to enhance phishing resilience

1.Gradient Boosting Classifier (GBC) - Gradient Boosting Classifier (GBC) is a powerful machine learning algorithm widely used for classification problems. It belongs to the family of ensemble methods, which combine multiple weak models to build a stronger predictive model. The core idea behind gradient boosting is to train models sequentially, where each new model focuses on correcting the errors made by the previous ones. Typically, decision trees are used as the base learners in GBC. GBC is favored because of its high predictive accuracy and flexibility. It can handle different types of data and loss functions, making it suitable for a wide range of applications such as fraud detection, phishing detection, and customer churn prediction. However, it can be computationally intensive and prone to overfitting if not properly tuned with parameters like

learning rate, number of estimators, and tree depth.Overall, Gradient Boosting Classifier is a robust and effective tool for building classification models with strong predictive power.

Analysis :

SVM Accuracy: 0.98

Accuracy is the overall correctness of the model. It means the model correctly predicted 98% of the cases (both positive and negative).

Confusion Matrix:
[[889 0]
[25 120]]

Precision Score: 1.0

Precision measures how many of the predicted positive cases are actually positive.

Precision = TP / (TP + FP) = 120 / (120 + 0) = 1.0

This means all the positive predictions made by the model were correct — no false positives.

Random Forest Accuracy: 0.98

Like before, accuracy of 0.98 means your model correctly predicted 98% of all cases.

Confusion Matrix:

[[888 1]
[21 124]]

Precision Score: 0.992

Precision = TP / (TP + FP) = 124 / (124 + 1) ≈ 0.992

This means about 99.2% of the time when your model predicted a positive, it was correct.

3. Naive Bayes Accuracy: 0.97

The model correctly predicted 97% of the total cases, which is slightly lower than SVM and Random Forest.

Confusion Matrix:

[[888 1]
[25 120]]

Precision Score: 0.9917

Precision = TP / (TP + FP) = 120 / (120 + 1) ≈ 0.9917

So about 99.17% of positive predictions were correct.

II. METHODOLOGY

In response to the growing sophistication of phishing attacks, we propose a comprehensive phishing detection system named PHISHNET. This system is designed to leverage advanced machine learning techniques and incorporate multiple data sources to enhance the accuracy and effectiveness of phishing detection across various communication channels, including URLs, emails. The proposed system aims to address key challenges such as personalization, cold start problems, and the temporal relevance of data.

A. EVALUATION OF MODEL

In this stage we evaluate the models by entering the dataset with predictor variables to each model, then the models will predict the targeting variable according to the prediction results and we will compare it with real values. In Fig. 2 there is a learning model of the proposed work.

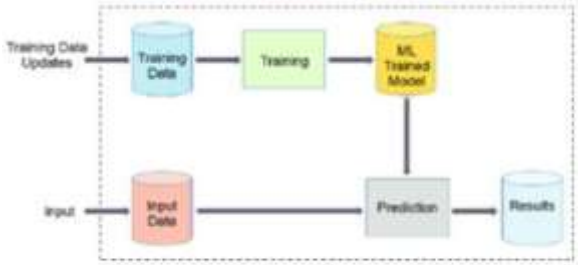

Fig. 2. Learning model from peposed system.

III. CONCLUSIONS AND FUTURE WORK

In this project, we developed and evaluated machine learning models for detecting phishing threats across two distinct channels: URLs, emails. The results demonstrated that our models achieved high accuracy rates approximately—98% for URL detection using Gradient Boosting, 97% for email detection with Random Forest. These findings underscore the effectiveness of machine learning techniques in identifying phishing attempts and enhancing
cybersecurity measures. The successful implementation of these models indicates that they can serve as valuable tools for organizations seeking to protect their users from phishing attacks. By leveraging advanced algorithms and comprehensive feature extraction methods, we have laid

the groundwork for a robust phishing detection system that can adapt to evolving threats. Looking ahead, there are several avenues for future work. First, we plan to expand the datasets used for training to include a broader range of phishing

examples, which could improve the models' generalizability and robustness. Incorporating more diverse features—such as behavioral analysis of users interacting with emails could also enhance detection capabilities. Additionally, exploring ensemble methods that combine multiple algorithms may yield even better performance by capitalizing on the strengths of different classifiers. Implementing real-time detection systems could further increase the practical applicability of our models, allowing organizations to respond swiftly to potential threats. Finally, ongoing research into adversarial machine learning techniques will be essential to ensure that our models remain effective against increasingly sophisticated phishing tactics. By continuously updating our models and methodologies in response to emerging threats, we can contribute to a safer digital environment for all users.

REFERENCES

[1] Adarsh Mandadi, Saikiran Bopanna, Vishnu Ravella, Dr. R Kavitha, Phishing Website Detection Using Machine Learning , 2022
IEEE 7th International conference for Convergence in Technology (I2CT).

[2] Edward Wijaya, Gracella Noveliora, Kharisma Dwi Utami , Rojali, Ghinaa Zain Nabiilah ,Spam Detection in Short Message
Service (SMS) Using Naïve Bayes, SVM, LSTM, and CNN , 2023 10th International Conference on Information Technology, Computer
and Electrical Engineering (ICITACEE) .

[3] Ammar Odeh , Ismail Keshta , Eman Abdelfattah, Machine Learning Techniques for Detection of Website Phishing: A Review for Promises and Challenges , 2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC).

[4] V Dharani, Divyashree Hegde, Mohana, Spam SMS (or) Email Detection and Classification using Machine Learning , 2023 5th
International Conference on Smart Systems and Inventive Technology (ICSSIT).

[5] Kerin Pithawala, Sakshi Jagtap and Preksha Cholachgud , Detecting Phishing of Short Uniform Resource Locators using
classification techniques , 2021 12th International Conference on Computing Communication and Networking Technologies
(ICCCNT).

[6] M.Rubin Julis et al. Spam Detection in SMS Using Machine Learning Through Text Mining, International journal of scientific & technology research, vol 9, Issue 02, 2020.