Transactions on
Federated Engineering & Systems
VOL. 1, ISSUE. 1, AUGUST 2023

RESEARCH ARTICLE Open Access

Phishing Websites Classification Placed on URL Features and Extreme Machine Learning

Ranjitha R G . Meenakshi Sundaram A

School of Computer Science and Engineering
REVA University, Bangalore, Karnataka, India.

Received: 21 September 2023 / Revised: 12 November 2023 / Accepted: 5 December 2023

DOI: **10.5281/zenodo.10279674**

Abstract – Phishing attacks have become an increasingly common threat to individuals and organizations alike. The traditional methods used to detect phishing attacks, such as blacklisting known phishing URLs or using heuristics to identify suspicious websites have proven to be limited in their effectiveness. Phishing attackers continuously evolve their tactics, making it difficult for traditional detection methods to keep up. To address this challenge, this study explores the use of machine learning classifiers to uncover illegitimate websites. Specifically, this research utilizes the Multilayer Perceptron and Bernoulli Naive Bayes (NB) classifiers. The feature selection process is performed using a decision tree classifier, which helps to identify the most relevant features for the classification task. To train and test the classifiers, the study collected a dataset of blacklisted and whitelisted websites. Accuracy, precision, recall, and the ROC curve were only few of the measures used to assess the classifier's effectiveness. The results demonstrate the effectiveness of the Multilayer Perceptron and Bernoulli NB classifiers in detecting phishing websites. The feed forward neural network classifier achieved an accuracy of over 82% on the dataset. These results showcase the potential of machine learning techniques in improving the discovering of phishing attacks and reducing further risks of phishing attacks.

Index Terms – Neural Network, Bernoulli Naive Bayes, phishing attacks, website security.

I. INTRODUCTION

Phishing is considered a cyber-attack in which the bad actor impersonates a trusty person or entity, such as a bank, social networking platform, or prominent e-commerce site. The attacker frequently uses email, messaging applications, or social media to trick the victim into divulging personal information like

login passwords, credit card details, or other sensitive data. Traditionally, individuals and organizations have relied on manual methods to detect phishing attacks, such as checking URLs or looking for misspellings and grammar errors in emails. However, these methods are not always reliable and may fail to detect well-crafted phishing emails. Additionally, attackers often employ tactics that make phishing emails appear to be legitimate, such as sending messages from email addresses that appear to be from trusted sources. As phishing attacks become more complex and sophisticated, it is increasingly essential to implement proactive and reliable methods for detecting and preventing them. Machine learning classifiers have emerged as a promising solution for detecting phishing attacks. These classifiers can identify patterns and features that distinguish phishing from legitimate websites and can detect and flag potential phishing websites.

By training classifiers on a dataset containing both phishing and legitimate websites, machine learning can identify distinct patterns and features that differentiate phishing from legitimate websites. These patterns can then be used by the classifiers to identify and flag potential phishing websites. In this context, the current study aims to explore the effectiveness of machine learning classifiers in uncovering phishing websites. Here the study utilizes the Feed Forward Neural Network alongside Bernoulli Naive Bayes (NB) classifiers, with feature selection performed using a decision tree classifier. The performance of the classifiers is assessed using variegated metrics, namely accuracy, recall, precision, precision recall graph and the Receiver Operating Characteristic (ROC) curve.

II. LITERATURE SURVEY

A comprehensive review of phishing detection techniques was conducted in [1] this study, the researchers collected and analysed a large amount of phishing data to better understand their prevalence, sources, and mitigating measures. To determine when and how quickly phishers act, domain names and site hosting were investigated. The study identifies at-risk service providers and locations that need improved phishing detection and prevention. APWG, PhishTank, OpenPhish, and Spamhaus collected three million phishing complaints from 2021 to 2022. This enabled a reliable dataset. Blacklists are often used to prevent phishing attacks. A blacklist's effectiveness leans on vastness, scale, restore frequency, accuracy, in addition to other considerations.[2] Compare Google Safe Browsing (GSB), OpenPhish and PhishTank phishing blacklists. These blacklists were also examined for adoption rate, dropout rate, lifespan, and URL overlap. GSB had a blend 1.6 million URLs, PhishTank had 12,433, and OpenPhish had 3,861. OpenPhish eliminated most of its URLs after 21 days, which may lessen the blacklist's efficacy. Phishing URLs tend to be transitory since their appearance in all three blacklists decreases with time. All three blacklists allow URL reuse. All three blacklists have seen several previously deleted URLs reappear within a day, suggesting premature removal.

Current anti-phishing approaches use external services and harmful site elements to detect phishing sites. These approaches time and skill to identify phishing characteristics. Phishing websites are not immediately detected when external services involved are used. [3] Phishing-website identification (RF) using a convolutional neural network and random forest is presented in this study. The technique can verify a URL's authenticity by reading its content or using other services. Character embedding transforms URLs into predetermined matrices, then CNN models extrapolates features within various

depths. [4] proposes an improved blacklist technique to phishing website detection. This method uses website source code to differentiate. The goal is to find fake clones of trusted websites with malicious code. Each phishing website is fingerprinted using the approved criteria. Each website in the database had a unique Simhash fingerprint. Filenames, pathnames, and tag attribute values are used to calculate a fingerprint. The tests found that 84.36% of phishing websites were copies with updated content. The suggested method, like the blacklist, may instantly detect cloned phishing sites.

III. DATASET

The dataset named "Phishing websites Data" is collected from Kaggle. It contains around 11,431 website URLs along with their class labels as phishing or legitimate. The features are extracted based on Address bar features, Abnormal page content features and Domain features such as:

- Long URL, (-) to domain, Double (@), Double (//), TinyURL shortening services, Page Rank, Domain age, DNS record, IP address, Subdomains, Domain Registration Length.
- The attribute values are integers 0 and 1, where 0 denotes legitimate and 1 denotes phishing website. The dataset is then split into two sections, 80% for training and 20% for testing.

IV. METHODOLOGY

A. Artificial Neural Network

Neural Network Classifier uses data patterns and correlations to classify fresh samples into predetermined categories. Labeled examples coaches the underlying classifier. The neural network learns to recognize class label patterns and characteristics in input data during training. Neural network classifiers include input, hidden, and output layers. The categorization model uses website attributes from the input layer. Implementation determines n. The model recognizes data patterns in hidden levels. 'm' and 'p' are the buried layer node counts. Rectified Linear Unit (ReLU) activation for hidden layers incorporates non-linearity and improves the model's capacity to grasp complicated patterns. The model's prediction of a phishing website is the output layer's single node. The model's prediction confidence is represented by the output layer's sigmoid activation function. A collection of categorized phishing and legal websites trains the neural network model. To reduce the discrepancy between projected and actual output, node connection weights are changed during training. The trained model can classify new websites as phishing or authentic depending on their attributes. Neural network models may find data patterns that people, or simpler models cannot. If the model is sophisticated or the dataset is limited, they may cause overfitting problem.

B. Bernoulli Naive Bayes Classifier

The Naive Bayes Classifier is a populus probabilistic measures-based machine learning algorithm utilized in variegated classification tasks including text classification and spam detection. In the context of phishing detection, the Naive Bayes classifier analyses a website's attributes and then calculates the likelihood that a website is a phishing website or a real website. The Naive Bayes classifier calculates the probability of website belonging to a particular class (legitimate or phishing) placed on its attributes. A vector of features represents each dataset website URL. On this dataset, Naive Bayes classifier is trained. Bayes theorem is used to compute the probability of each characteristic given the class (phishing or

genuine) during training. Using training data, probabilities are generated to develop a model that can categorize new websites as phishing ones or authentic based off of their properties. The Naive Bayes classifier analyses the probability of each characteristic given the class and multiplies them to get the website's class. Next, the algorithm analyses the website's the probability of being a phishing site versus its probability of being a valid site and assigns it to the class with the higher probability.

C. Proposed System

The approach proposed to handle phishing attacks is to utilize the Uniform Resource Locator (URL) of a website in detection mechanism as it specifies where a resource on the internet may be found. The first step in this approach entails collecting a dataset consisting of both legitimate and phishing websites. The URLs are subsequently retrieved from the dataset using feature extractor to capture relevant features which can help distinguish the legitimate and phishing sites. These features include URL length, number of subdomains, domain age shortening services, domain registration length, DNS record, usage of suspicious URL symbols and Page rank. The decision tree algorithm builds a tree-like structure, where in every one of the nodes portrays a feature, and each branch portrays a probable value of the specific feature traversing through all feature values. The data is then parted as training and testing sets. The training set is used to train the classifiers on the relevant features selected by the decision tree algorithm. The testing set is utilised to estimate the performance of both classifiers. The execution of both classifiers is measured with metrics namely accuracy, precision, recall, and F1-score. Accuracy measures overall performance of the classifiers, while precision and recall measure the performance of the classifier in identifying phishing and legitimate websites, respectively. The evaluation model shows that the neural network classifier outperforms the naive Bayes classifier in terms of accuracy, precision, recall, and F1-score.

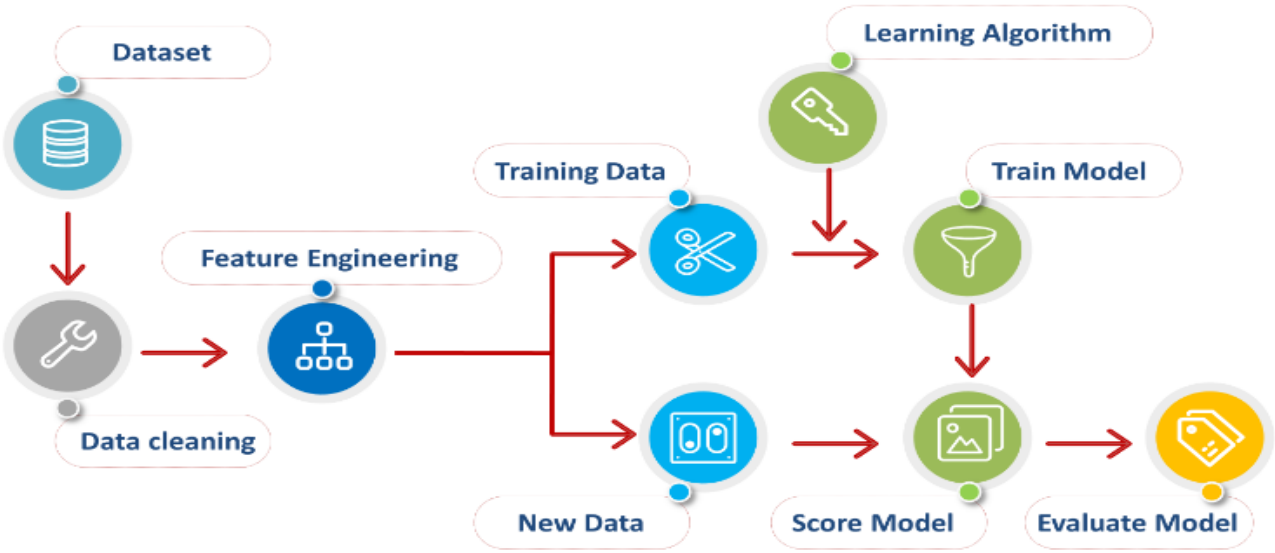

Fig.1 Architecture of the Proposed System

V. RESULTS AND DISCUSSION

Confusion Matrix of Naive Bayes and Neural Network Models:

The confusion matrix summarizes the performance of a classification of both the models where it is typically divided into four quadrants: True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) given below in Fig 2 and Fig 3.

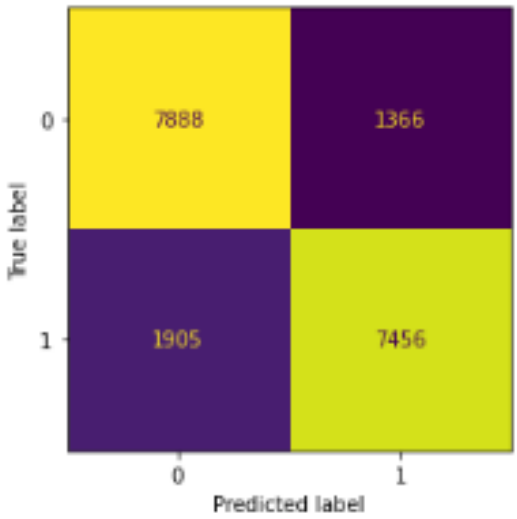

Fig 2. Neural Network Confusion Matrix

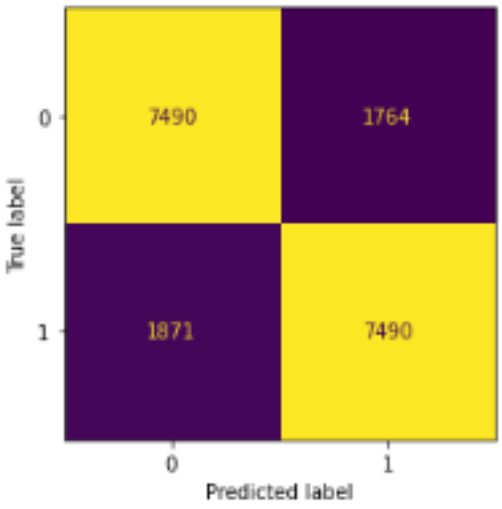

Fig 3. Naive Bayes Confusion Matrix

ROC Graph: The ROC graph shows the model's performance across all thresholds by showing the TPR and FPR for each threshold. A classifier with a TPR of 1 and FPR of 0 correctly identifies all positive samples and none of the negative ones. The ROC graph is a good way to evaluate binary classification models since it shows performance across all threshold values, compared to accuracy, precision, and recall.

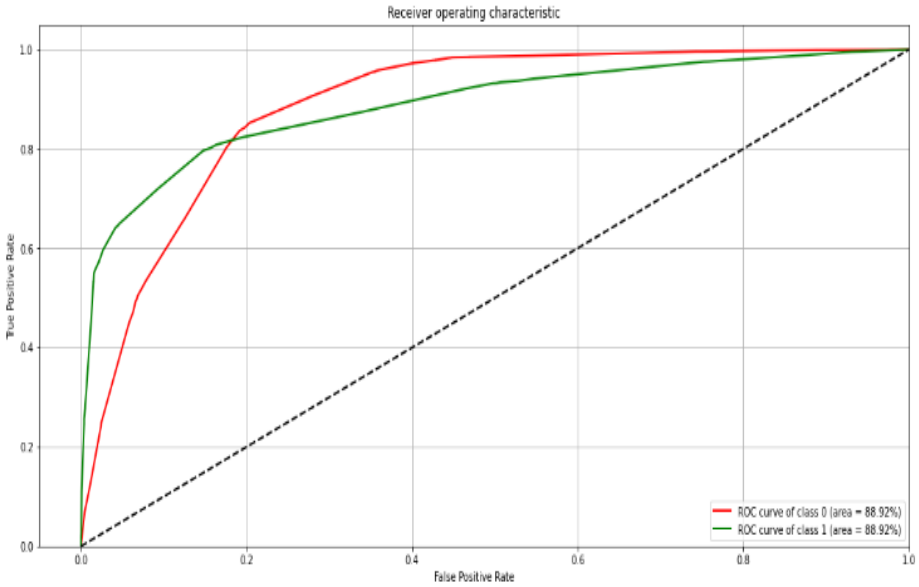

Fig 4: Receiver Operating Characteristic Curve

VI. CONCLUSION

The evaluation of both machine learning models showed Accuracy to be one of the crucial metrics for assessing the performance of such models. The Neural Network Classifier had an Accuracy of 82%, while the Naive Bayes Classifier had an Accuracy of 80%. Although the Neural Network Classifier outperformed the Naive Bayes Classifier slightly, the difference in Accuracy is relatively small, and both models showed promising results. Overall, machine learning methods like the Neural Network Classifier and Naive Bayes Classifier can detect phishing website attributes. As attacker's strategies evolve, newer phishing detection and prevention solutions must be developed and tested.

REFERENCES

1. Aaron, G., Chapin, L., Piscitello, D., & Strutt, C. (2020). Phishing Landscape 2020: A Study of the Scope and Distribution of Phishing.
2. Yang, R., Zheng, K., Wu, B., Wu, C., & Wang, X. (2021). Phishing website detection based on deep convolutional neural network and random forest ensemble learning. *Sensors*, *21*(24), 8281.
3. Ahmed, S. T., Sreedhar Kumar, S., Anusha, B., Bhumika, P., Gunashree, M., & Ishwarya, B. (2020). A generalized study on data mining and clustering algorithms. *New Trends in Computational Vision and Bio-inspired Computing: Selected works presented at the ICCVBIC 2018, Coimbatore, India*, 1121-1129.
4. Rao, R. S., & Pais, A. R. (2017). An enhanced blacklist method to detect phishing websites. In *Information Systems Security: 13th International Conference, ICISS 2017, Mumbai, India, December 16-20, 2017, Proceedings 13* (pp. 323-333). Springer International Publishing.
5. Saha, I., Sarma, D., Chakma, R. J., Alam, M. N., Sultana, A., & Hossain, S. (2020, August). Phishing attacks detection using deep learning approach. In *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)* (pp. 1180-1185). IEEE.
6. Ragaventhiran, J., Vigneshwaran, P., Kodabagi, M. M., Ahmed, S. T., Ramadoss, P., & Megantoro, P. (2022). An unsupervised malware detection system for windows based system call sequences. *Malaysian Journal of Computer Science*, 79-92.
7. Wu, L., Du, X., & Wu, J. (2015). Effective defense schemes for phishing attacks on mobile computing platforms. *IEEE Transactions on Vehicular Technology*, *65*(8), 6678-6691.

8. Sreedhar, K. S., Ahmed, S. T., & Sreejesh, G. (2022, June). An Improved Technique to Identify Fake News on Social Media Network using Supervised Machine Learning Concepts. In *2022 IEEE World Conference on Applied Intelligence and Computing (AIC)* (pp. 652-658). IEEE.
9. Benavides, E., Fuertes, W., Sanchez, S., & Sanchez, M. (2020). Classification of phishing attack solutions by employing deep learning techniques: A systematic literature review. *Developments and Advances in Defense and Security: Proceedings of MICRADS 2019*, 51-64.
10. AlEroud, A., & Karabatis, G. (2020, March). Bypassing detection of URL-based phishing attacks using generative adversarial deep neural networks. In *Proceedings of the sixth international workshop on security and privacy analytics* (pp. 53-60).
11. Muppavarapu, V., Rajendran, A., & Vasudevan, S. K. (2018). Phishing detection using RDF and random forests. *Int. Arab J. Inf. Technol.*, *15*(5), 817-824.