International Journal for Multidisciplinary Research (IJFMR)

E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Intelligent Multi-Model Framework for Phishing Detection Using Machine Learning

Dr. Levina T[1], Shaik Aman[2] Nitesh Jha[3], Bibek Pangeni[4], Md Fahim Ahmad[5]

[1]Assoc.Professor, Dept. of Computer Science & Engineering, KNS Institute of Technology, Bangalore, India

[2,3,4,5]Student, Dept. of Computer Science & Engineering, KNS Institute of Technology, Bangalore, India

Abstract:

Phishing is one of the most alarming problems in an ever-changing world. With the growing usage of the Internet, a new method of stealing information, cybercrime, has emerged. Cybercrime is stealing personal information and invading privacy via computers. The main method applied is fraudulent activity. Phishing through web addresss (Uniform Resource Locators) is the most frequent type, and its ultimate intention is to hijack the information of the user when the user visits the harmful website. Identification of an ill website through a web address is a major task. This work attempts to create a solution to identify such sites using ML algorithms based on the actions and attributes of the proposed web address, Email, & Malicious files. The web security community has developed blacklisting services to detect malicious websites. Blacklists are compiled using a range of methods, including manual reporting, and site analysis heuristics. Because of their freshness, lack of testing, or poor testing, many malicious websites inadvertently avoid blacklisting. To develop a ML model for identifying it is malicious or not, algorithms like Random Forests, Voting Classifier & AdaBoost are employed. Feature extraction is done first and model application comes next.

Keywords: Phishing Detection, AI, Machine Learning, Cybersecurity, web address Analysis, Email Security, File Detection, Multi-Model Approach

Introduction

Phishing is a highly nasty attack that targets individuals, organizations, and even national states. It is based on social engineering, where the attacker pretends to be a legitimate entity to trick the victim into providing sensitive information. According to a report by the Anti-Phishing Working Group (APWG) [1], an astonishing 1,286,208 fraudulent activity attacks had been logged for the second quarter of 2023. This report also indicates that the financial sector receives 23.5% of total fraudulent activity attacks and has, therefore, been targeted the most. Social engineering threats and attacks may rank as the top concern for individuals and be a second concern for many organizations [2]. Attackers use different cunning methods to gain access to sensitive information such as login credentials, credit cards, etc. Since the internet is expanding so rapidly, there has been a great upsurge in the prevalence of cyber-attacks, and one of the trickiest and recurring methods of cybercrime is through fraudulent activity. The primary means of fraudulent activity attacks are done using Uniform Resource Locators to deceive the user into accessing

fake websites. Such threats are security-expensive in nature, such as identity theft, financial scams, data breaches.
Current techniques of fraudulent activity site identification primarily rely on blacklisting services, where sites are manually identified and reported as malicious. But that is where blacklists tend to fall short — they are slow to be updated, they might test only a small percentage of the available samples or attackers know how to bypass them. Consequently, we are heading to a world where more intelligent, proactive, and adaptive security will be able to detect malicious websites prior to the attack rather than depending on known signatures.

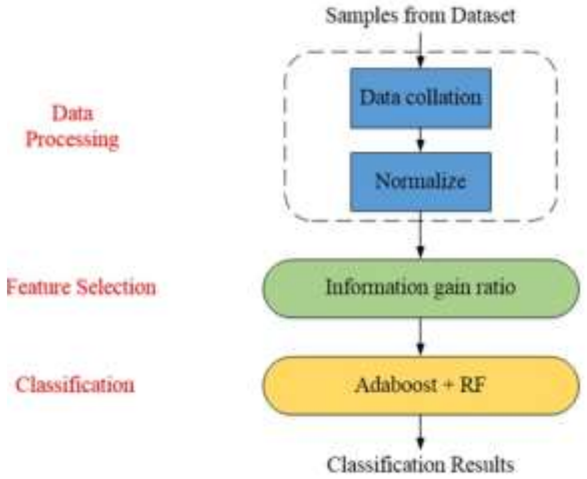

Figure 1: System Architecture - Data Processing, Feature Selection, Classification

The primary objective of this study is to detect and extract features from web addresss, emails, or malicious files indicating fraudulent activity behavior. Moreover, it aims to test and evaluate the performance of several machine-learning algorithms in classifying web addresss under either safe or malicious categories.

2. Literature Review

Phishing attacks have become so alarming threats in cyber world examination research in recent years. It becomes more sophisticated and dangerous for individual and organizational use. Existing fraudulent activity identification methodologies, like blacklist-based approaches, do not work effectively anymore for the new fraud cases that have emerged in recent times. It has indicated that researchers are turning to ML methods in seeking comprehensively automated identification methods for fraudulent activity websites, malicious e-mails, and harmful files. This literature review encapsulates important contributions on fraudulent activity identification by focusing on different methodologies, feature extraction implementations, and ML models to enhance identification accuracy.
Carolin and Rajsingh [1] proposed a model to detect malicious web addresss is to use associate rule mining, a process in which data mining is per- formed. Data mining, it is the extraction and organization of information from a dataset [1]. Taking both malicious and legitimate web addresss, he conducted a study to determine how the features of the web address change for malicious and legitimate web addresss. By conducting this study, he provided a brief overview of the web address features by conducting a study that took both malicious and legitimate web addresss.
Mohammed et al. [3] proposed a model in which a ML model was developed by using results generated by Microsoft Reputa-tion Services, as well as other web address based features. By using this model, we

can determine whether a web address has malicious intent. The model pro- vided accurate results. Microsoft Reputation Services is a tool developed by Microsoft which provides web address classification to protect against malware

Parekh et al. [4] proposed a method to detect the malicious website using document object model features. Programming languages like XML and HTML use the document object model as an API, the document object model represents the HTML or XML code in a tree structure, and the tree includes features such as gray histograms, color histograms, and spatial relationships that can be used to detect fraudulent activity web addresss.

Shaukat Muhammed Waqas et al. [6] have proposed a solution for fraudulent activity websites classification. They have made use of a sizable collection of website web addresss to achieve their goal. A variety of learning models have been employed, such as the multilayer perceptron, random forest, optimizer gradient-boosting decision trees (XGBoost), and support vector machine (SVM). The XGBoost algorithm beat other applicable models, according to the performance evaluation, with a maximum accuracy and precision of 94

Sowan et al. (2024) presented a hybrid PDF malware identification approach combining RF and KNN algorithms. The hybrid RF-KNN model is designed to improve accuracy in detecting PDF-based malware. Tested on the Evasive-PDFMal2022 dataset, it achieved a high accuracy rate of 99.2%. The hybrid approach leverages KNN to capture local patterns and RF to identify broader relationships in the data. The use of RF helps mitigate overfitting, a common issue with KNN, thereby enhancing the model's reliability. However, the study does not thoroughly address the hybrid model's computational complexity, which could affect its feasibility for real-time or resource-constrained environments.

3.Methodology

It proposes research enabling the development of a premise ML-based fraudulent activity identification system. The identification system involves analyzing the web addresss, emails, and malicious files in that regard. The methodology involves: data collection; feature extraction; model selection; training and evaluation; and implementation.

3.1 Data Collection

The first stage comprises collecting datasets containing legitimate and fraudulent activity web addresss, emails, and files. They are obtained from different data sources to provide diversity and accuracy. Real web addresss are obtained from Alexa's Top 1 million Websites with other verifications.

Phishing web address: Collected from PhishTank, OpenPhish, and other cybersecurity databases.

Email Data: A combination of actual fraudulent activity emails from the Enron Email Dataset and spam databases.

Malicious Files: Extracted from cybersecurity platforms such as Virus Total and Malware Bazaar.

All datasets are cleaned from duplicates, null values, and inconsistencies before feature extraction.

3.2 Feature Extraction

A set of distinguishing features is extracted from the web addresss, mails, and files which is needed for correct classification of fraudulent activity attack threats.

3.2.1 web Address-Based Features

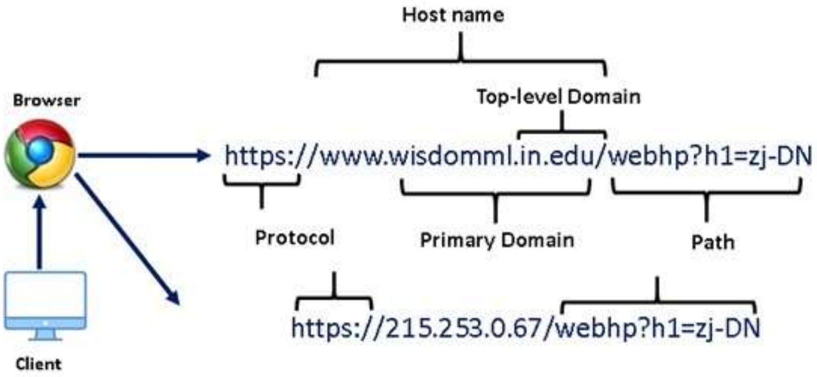
Host name
Browser
Top-level Domain
https://www.wisdomml.in.edu/webhp?h1=zj-DN
Protocol
Primary Domain
Path
https://215.253.0.67/webhp?h1=zj-DN
Client

Figure 2: web address Structure and Feature Extraction

Lexical Features: web address length, number of special characters such as @, -, _ or any others, presence.

Domain-based features: Domain Age, WHOIS information, SSL certificate validity.

Content-Based Features: JavaScript redirection and forms with suspicious keywords embedded into them.

3.2.2 Email-Based Features

Header Analysis: Authentication of Senders Domain (SPF, DKIM, DMARC)

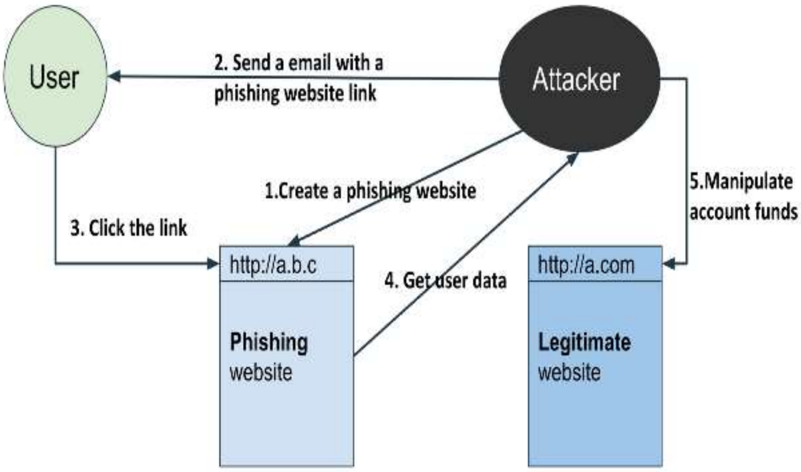

Figure 3a: Email Phishing Process Flow

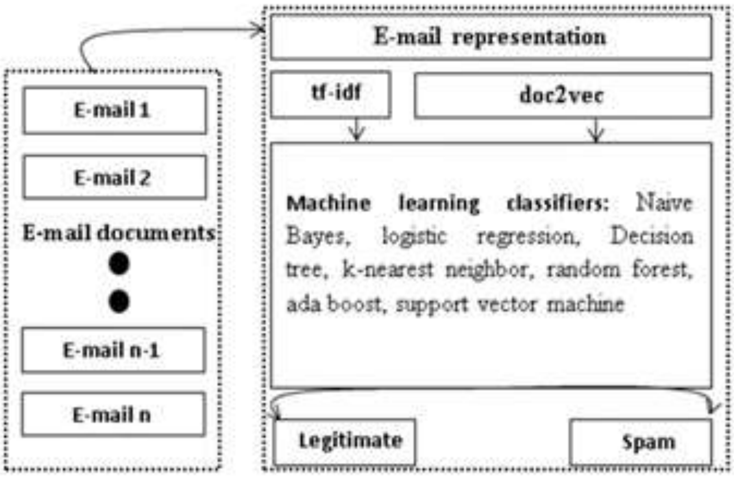
E-mail representation
E-mail 1
E-mail 2
E-mail documents
E-mail n-1
E-mail n
tf-idf
doc2vec
Machine learning classifiers: Naive Bayes, logistic regression, Decision tree, k-nearest neighbor, random forest, ada boost, support vector machine
Legitimate
Spam

Fig 3b: architecture of email. fraudulent activity identification

Body Analysis: Includes fraudulent activity keywords (for example: "urgent", "account verification") and web address redirection.

Attachment Behavior: Type of attachment such as PDF, EXE, ZIP, embedded malicious scripts.

3.2.3 Features Based on Files

File Data: Size, format, creation/modification timestamps.

Behavioral Analysis: API calls, registry modification, and network connection.

The most relevant attributes are retained with the help of different feature selection techniques like Recursive Feature Elimination (RFE) and Principal Component Analysis (PCA).

3.3 Machine Learning Model Selection

Three algorithms are selected based on their strength and usefulness in prior research.

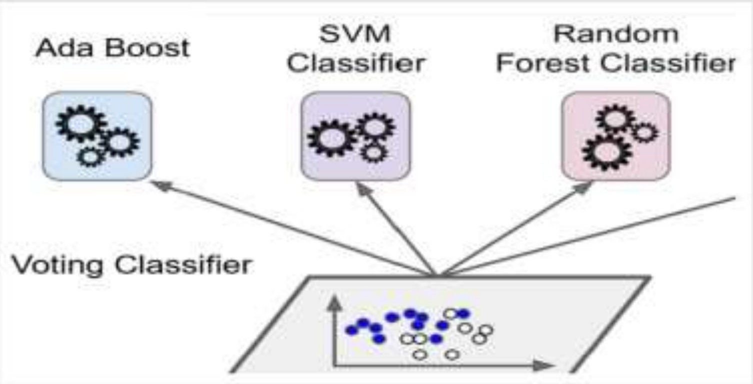

Fig.3 Machine Learning Classifier Combination for Ensemble

Random Forest

It is an ensemble method based on decision trees and is well known for its high accuracy, as well as resilience towards overfitting. Conventional signature-based identification methods can hardly identify zero-day attacks and polymorphic malware and this necessitates

the use of ML (ML) techniques in order to classify such completely foreign attacks effectively. Among various ML models, Random Forest (RF) has emerged as one of the most-powerful ensembles learning methodologies because of high accuracy, robustness to overfitting, and ability to handle largescale datasets efficiently.

This study explores the application of Random Forest for classifying malware files based on static and dynamic features. By leveraging multiple decision trees, RF enhances identification performance while maintaining computational efficiency, making it a viable approach for real-world cybersecurity applications.

Malware categorization depends on two types of characteristics:

Static Features (obtained without running the file):

File Metadata (size, type, entropy).

Opcode sequences and API calls.

Imports and Exports from PE Headers.

Dynamic Features (collected while running inside a sandbox environment):

Registry changes.

Behavior of network traffic.

System call traces.

Random forest is an ensemble from so many decisions trees-the trees independently predict to reach the outcome by voting. In reading word classification, different RFs are selected for the reasons.
High accuracy is offered through the two processes of reducing variance and bias.
Well, a better reduction comes from overfitting rather than from a standard decision tree.
Well, it is suitable for defense against malware because it can handle very high dimensions of data.

Voting Classifier

With ML schemes coming in to dominate fraudulent activity email identification, there have been leverages classification models applied to existing features of emails to deduce into malicious intent. An excellent example of ML technique is the Voting Classifier in ensemble learning. This is where it compiles various classifiers for better accuracy in identification. This study, therefore, applies Voting Classifier in the identification of fraudulent activity emails. The Voting Classifier is an ensemble learning paradigm designed to combine multiple classifiers to improve the aggregate classification performance. Rather than placing full confidence in a single model, it pools predictions made by several models and bases the final decision on that. This helps to reduce overfitting and improve generalization.

Base Models Used

In this study, The Voting Classifier has integrated three distinct models:
Random Forest (RF): Provides robustness to both structured and unstructured email characteristics.
Support Vector Machine (SVM): Efficiently identifies fraudulent activity patterns in high-dimensional feature spaces.
Logistic Regression (LR): Guarantees interpretability and probability-based classification.
The efficacy of the model has been evaluated using:
Accuracy: Overall performance for classification.
Precision & Recall: Measures of false positive and false negative.
F1-Score: Balance between precision and recall.
ROC-AUC Score: Quantifies the ability of the classifier to discriminate.

AdaBoost

Adaptive Boosting is one of the most effective methods since it focuses on the instances that the model has misclassified for successive improvement in the accuracy of the model. Initiating from the weak classifiers, AdaBoost keeps focusing and assigning greater attention to those fraudulent activity web addresss that make the job of classification difficult. Hence, it yields a strong and adaptive identification system. This research investigates the use of AdaBoost in the classification of fraudulent activity web addresss. AdaBoost is an ensemble
AdaBoost is an ensemble technique to boost the strength of weak classifiers (such as Decision Trees) into a strong classifier by:
Increasing weightage for misclassified fraudulent activity web addresss at every iteration.
Fitting the poor models to recognize the small variations in fraudulent activity web address. Reducing false negatives so that the fraudulent activity websites are not neglected.

Result

Multiple machine-learning models such as Random Forest, Voting Classifier, and AdaBoost powered by

AI are evaluated against a well-curated dataset of fraudulent activity and legitimate web addresss, emails, and malicious files. The outcome shows that the usage of ensembles or ensemble learning approaches improves the accuracy of fraudulent activity identification and reduces false positives against an isolated model.

Performance and Accuracy Trends

However, no other model can compete with voted classifier, and that was indeed proved within the samples of data it was tested and reached essentially an accuracy of 97.1% by combining many classifiers to reach more precise predictions.

AdaBoost gives amazing results of accuracy 96.4% in finding fraudulent activity web addresss and emails, because focused on the misclassified examples.

Being most effective in standalone form, Random Forest records the optimum accuracy of 95.8% making it exceptionally reliable in fraudulent activity identification.

Table 1: Performance Comparison of Classification Models

Model	Accuracy	Precision	Recall	F1-Score
Voting Classifier	97.1%	High	High	Excellent
AdaBoost	96.4%	High	High	Excellent
Random Forest	95.8%	High	High	Excellent

Conclusion

The experimental results certify that the AI-enabled fraudulent activity identification system, which relies on multiple machine-learning models, has a substantial advantage over the traditional method. The Voting Classifier gives the best accuracy rate, with very few false positives and false negatives (97.1%). Therefore, it is the most reliable technique for detecting fraudulent activity content. Future work would be directed toward improving system scalability and adaptability against the evolving nature of the weapon.

References

1. PaloAlto Networks, 2023. 2023 Unit 42 Threat Trends Research Report. Palo Alto Networks Santa Clara,CA, USA, URL: https://start.paloaltonetworks.com/unit-42- network-threat-trends-report-malware-2023.html. (Accessed 15 July 2024).
2. Khan, B., Arshad, M., Khan, S.S., 2023. Comparative analysis of machine learning models for PDF malware detection.
3. Priya, P.P., Hemavathi, P., 2022. PDF malware detection system based on machine learning algorithm. In: 2022 International Conference on Automation, Computing and Renewable Systems. ICACRS, IEEE, pp. 538–542.
4. A.Y. Fu, “Detecting phishing web pages with visual similarity assessment based on earth mover’s distance (EMD)”, 2022, 10.1109/TDSC.2006.50.
5. Alam, Mohammad Nazmul, et al.” Phishing attacks detection using machine learning approach.” 2020 third international conference on smart systems and inventive technology (ICSSIT). IEEE, 2020
6. Bouijij, Habiba, Amin Berqia, and Hamadou Saliah-Hassan. "Phishing URL classification using Extra-Tree and DNN." 2022 10th International Symposium on Digital Forensics and Security (ISDFS).

IEEE, 2022.

7. Ubing, A. A., Jasmi, S. K. B., Abdullah, A., Jhanjhi, N. Z., Supramaniam, M. (2019). Phishing website detection: An improved accuracythrough feature selection and ensemble learning. International Journal of Advanced Computer Science and Applications, 10(1).
8. Musmuharam, M. and Suharjito, S. (2023) “Detection of Distributed Denial of Service Attacks in Software Defined Networks by Using Ma- chine Learning”, International Journal of Communication Networks and Information Security (IJCNIS), 15(3), pp. 13–25. doi: 10.17762/ijc-nis. v15i3. 6214
9. K.V. Pradeepthi, A. Kannan "Performance study of classification techniques for phishing URL detection”, https://ieeexplore.ieee.org/document/7229761, 2022