IRASS Journal of Multidisciplinary Studies
https://irasspublisher.com/journal-details/IRASSJMS
ISSN *(Online) 3049-0073*

Phishing Website Detection through Machine Learning Algorithms: A Comparative Analysis

Ochuko Piserchia*

Independent Researcher

Corresponding Author: **Ochuko Piserchia**

Independent Researcher

Article History

Received: 06 / 10 / 2025

Accepted: 29 / 11 / 2025

Published: 08 / 12 /2025

Abstract: Phishing is the attempt to acquire sensitive information, often for malicious reasons, by masking as a trustworthy entity in an electronic communication. Once victims access a phishing website, the attacker attempts to convince them to send their private information such as usernames, passwords and credit card resulting in information theft.

Despite the growing awareness of phishing and its prevention through traditional methods such as DNS filtering, blacklisting, and user awareness trainings regarding the problem and its associated risks, it remains as growing concern, costing millions of dollars each year. The only effective defense against these threats is accurate detection of phishing attempts. However, machine learning methods have shown reasonable performance rates. Machine learning techniques which are a subset of artificial learning (AI) have shown significant success in detecting phishing websites in comparison to traditional methods, although effectiveness can vary depending on the approach deployed.

This research aimed to solve this problem by analyzing a phishing website dataset with six supervised algorithms. This was achieved using a feature selection investigation on the most promising of the 6 algorithms using primarily the filter method and compared with outcome of wrapper method. In addition to Accuracy and ROC (Receiver Operating Characteristic) Curve performance metrics, we also considered MCC (Matthews Correlation Coefficient). The experiment showed that Random Forest is the best performing algorithm at 0.989 MCC score (97% accuracy). We also realized 5 of the 30 features are enough for the classification with little or no reduction in performance.

Keywords: *Phishing Detection, Machine Learning, Comparative Analysis, Random Forest, Feature Selection, Cybersecurity.*

How to Cite in APA format: Piserchia, O. (2025). Phishing Website Detection through Machine Learning Algorithms: A Comparative Analysis. *IRASS Journal of Multidisciplinary Studies, 2(12), 11-29.*

Chapter 1

Introduction

Phishing attacks involve the unauthorized access or acquisition of sensitive information electronically using deceptive techniques and social engineering tactics. With the vast availability of data in the digital ecosystem and increased dependance on online platforms, phishing attacks continue to pose a significant threat to users' safety. One of the most predominant forms of phishing is through phishing websites. This occurs when attackers camouflage malicious phishing websites to look legitimate with the intent to trick users to visit them. When this happens, they gain access to sensitive and personal information of users.

Phishing attacks continue to gain traction and poses a huge threat to individual users and businesses across the globe. Despite the knowledge of this threat, the risk of falling victim to phishing attacks have increased as the attackers are also deploying innovation ways to invade and outsmart cyber security systems. These attackers often utilize fraudulent URLs, emails or messages that appear to originate from reputable entities, tricking unsuspecting users into sharing their confidential data. Machine Learning (ML) techniques have shown promise in detecting phishing websites, although the outcome varies depending on the technique deployed. To this end, this research paper aims to is to compare the effectiveness of ML classification models in detecting phishing websites.

Background of the Study

Phishing is believed to have first been described in an article by Felix, Jerry, and Hauck (1987) titled "System Security: A Hacker's Perspective" but it took roughly a decade after increased has with electronic communication and internet. Despite the awareness of phishing among internet users, the problem remains as potent as ever. claiming millions of dollars annually. According to a study conducted by the Anti-Phishing Working Group, in 2017, more than 291,000 different phishing websites detected. Over 592,000 unique phishing email campaigns reported, and more than 108,000 domain names attacked. An article on Forbes website in May 2017 claims Phishing Scams cost American Businesses half a billion dollars each year.

As stated already correct detection of phishing attempts is the surest form of defense, there are other non-technical solutions put in place to address this problem including legal and education solution. For example: Followed by many countries, the United States was the first to enact laws against phishing activities and many phishers have been arrested and sued. Phishing has been

added to the computer crime list for the first time on January 2004 by "Federal Trade Commission" "FTC" which is a U.S government agency that aims to promote consumer protection.

Though the awareness of phishing has grown over time, the technical knowledge required to identify potential phishing website remains untapped. Most internet-users lack basic knowledge of current online threats that may target them. They also find it difficult to differentiate how legitimate online sites formally contact their consumers in case of an information update or maintenance. This makes potential victims ignore security indicators that should have aroused their suspicion and instead follow the prompts of the attacker.

A major form of defense is user awareness of this attack and continues deployment of innovative measures to avoid them. On of such defensive measures is Machine Learning. (ML). According to Burlela (2023), ML has emerged as a promising approach in detecting phishing. ML uses data and algorithms to train machines to think like the human brain, learn from experiences and identify patterns. This technology allows computers to improve their performance based on pattern recognition from past experiences without being explicitly programmed for each task. The ability for ML algorithms to recognize patterns offers a breakthrough to detect and classify phishing attacks by analyzing patterns and indicators of fraudulent activity based on historical data. Leveraging ML models enhances detection capabilities and accurately predict whether a webpage is a phishing site or legitimate.

The objective of this research paper is to compare the effectiveness of ML classification models in detecting phishing websites. By identifying the most accurate ML model among the considered algorithms, the aim is to enhance detection capabilities and mitigate the risks associated with visiting phishing websites, ultimately restoring consumer trust.

To address the effectiveness of ML algorithms in detecting phishing websites, this research will address two questions including: "How effectively are ML algorithms detecting phishing websites targeting users"? and "What machine learning techniques are most used to identify phishing websites, and how they perform in the context of cybersecurity threats"? Addressing these questions will enhance online security and safeguard user information which is critical to build trust in businesses.

Problem Statement

The rise of e- commerce have led to major reliance on internet and digital services for daily transactions such as internet/online banking, social networking, online shopping, accessing education resources, booking hospital appointments etc. These increase internet-based services have in turn lead to increase in cyber threats such as Phishing leading millions in financial losses globally. According to the Global State of Fraud and Identity Report (2024), 80% of organizations across the global experience payment frauds due to attacked from phishing websites while 20% of customers have reported being victims of online fraud. In an article by Chisom (2024), the four largest economies in Africa- South Africa, Egypt, Nigeria and Kenya suffered a cumulative loss of over Two million dollars to phishing related crimes. This has a significant impact to the growing economies.

According to a Harvard Business Review by Isik and Goswami (2024), there was a 60% increase in phishing attacks due to AI deepfakes. To safeguard users, several attempts have been made to create awareness of phishing, yet the risk remains eminent. The effectiveness of these attempts has been limited by incompleteness of the list of features used and misuse of performance metrics. Relevant research for literature to support this also did not produce a clear overview of all the major approaches in this area. A major challenge is that users still struggle to differentiate the difference between legitimate and phishing website. A collaborative effort that captures the techniques, data sets, and algorithms used in phishing website detection was not available in a methodical format. Therefore, there is a need to study this area to provide a holistic overview to tackling the problem with phishing websites.

ML algorithms offer a promising solution of detecting phishing websites when compared to traditional approaches. Due to the artificial intelligence (AI) nature of ML algorithms, they quickly detect phishing websites by analyzing various features of the webpages such as URL structure, domain characteristics, and content analysis. Though ML algorithms have demonstrated significant progress in detecting phishing websites, the results varies based on the algorithm deployed. This has given rise to the need to conduct a comprehensive comparative analysis of six ML algorithms to identify the most accurate and efficient models.

Several collaborative research efforts have gone into the use of ML models to detect phishing websites. Mohammad et al. (2012) proposed a rule-based data mining classification techniques using 17 different features to distinguish phishing from legitimate websites. Abdelhamid et al. (2013) introduced the Multi-Label Classifier based Associative Classification (MCAC) to detect phishing websites Following their study, Mohammad et al. (2014) developed a smarter model to enhance accuracy in predicting phishing attacks based on self-structuring neural networks. All of these collaborative research is an addition to other existing ML algorithms such as, Neural network (NN), Support vector machine, (SVM), Naïve Bayes (NB), and other ML classification techniques.

There are several traditional techniques that have been employed to detect and curb the risk of phishing websites, though the accuracy of these attempts were not impressive. This has led to many legitimate websites being classified as phishing Ali (2017). There are notably two traditional approaches to detect phishing websites. The blacklist and whitelist-based approaches depends on the blacklist or whitelist to verify of the currently visited website is either a phishing or legitimate website. The shortfall of the blacklist and whitelist based approach is that it cannot distinguish the newly created phishing websites from websites. Unlike traditional approaches, ML algorithms are trained to detect phishing websites by analyzing the ley features of the websites such domain name, URLs, context making it a smarter and more effective approach to resolving the problem with phishing.

Research Questions

This research will evaluate and respond to the following key questions:

1. What machine learning techniques are most used to identify phishing websites, and how the identified ML algorithms perform in regard to identifying phishing websites in the context of cybersecurity threats?

Machine Learning (ML) Algorithms

According to Gresele (2023), ML algorithms have become increasingly effective in detecting phishing websites, offering smart, adaptive and proactive defenses against evolving cyber threats. As mentioned previously, traditional methods like blacklist and whitelist approaches, and DNS filtering often fall short due to their dependance on known malicious URLs, which can quickly become dated as attackers become more innovative in creating new domain names. ML models are trained to analyze patterns and features of URLs and web pages, enabling them to identify new and unknown phishing attacks, even if they have not been previously encountered.

Machine learning focuses on developing algorithms that reason and think like a human brain and generate patterns and rules from past data and external supplied instances to develop models that are able to make predictions about future occurrences. ML algorithms are trained using datasets to learn from past experiences and improve their accuracy and performance with time. This makes them more effective at detecting phishing websites. The ability for ML algorithms to analyze various features of a website such as its context and web pages makes it more effective in detecting phishing websites by simply analyzing different URLs or domains. The use of ML algorithms for detection of phishing websites is a more proactive, adaptive and effective approach in guard railing users against evolving phishing attacks. Detecting phishing website is a critical step necessary to prevent phishing attacks. This research will explore the effectiveness of the following ML algorithms or models as the solutions to the problem of phishing websites.

Supervised Machine Learning

ML algorithms are classified as supervised learning when datasets are trained on known labels with instances in the training stage. Due to the training of the datasets, the model can classify new websites as either phishing or legitimate based on the patterns identified and learning that occurred during training and from past experiences. One of the advantages of supervised ML techniques is their ability to detect phishing in real time, helping organizations stay ahead of the problem in mitigating the risks even before it occurs. For instance, ML algorithms such as Support Vector Machines (SVM), Decision Tree (DT/J48), Naïve Bayes (NB), Random Forest (RF), Neutral Networks (NN), and Logistics Regression (LR) can improve their accuracy in detecting phishing by training input datasets, thus reducing the risk of false positives and false negatives.

Support Vector Machines (SVM/SMO)

SVM is a supervised ML algorithm typically used for regression and classification tasks. It is used in carrying tasks such as image classification, biometrics informatics, text categorization and effective in high dimensional spaces. SVM is highly effective in detecting phishing websites because of their ability to classify patterns and handle high dimensional data. Phishing detection involves analyzing various features of a website (e.g., URL structure, domain information, page content) to determine whether it is legitimate or malicious. The goal of an SVM is to find the optimal hyperplane that best separates the data points of different classes in the feature space. The hyperplane is chosen to maximize the margin, which is the distance between the hyperplane and the nearest data points from each class, known as **support vectors** (Cortes & Vapnik, 1995). However, recent advancements in SVMs combines hybrid SVM models with deep learning techniques, improved kernel functions, and applications in big data analytics (Zhang et al., 2021). Researchers have also explored SVMs for imbalanced datasets and multi-class classification problems (Wang et al., 2021).

Decision Trees (DT)

Just like SVM, DT/J48 is a supervised ML algorithm used for both classification and regression tasks. It models decisions and their possible outcomes in a tree-like structure where the internal nodes represent decisions based on features such as URL or text features. The branches represent the outcome of a decision while the leaf nodes represent the data classes or label ("Phishing" or "Legitimate"). Decision trees can handle numerical and categorical data easy. They are instinctive, and easy to interpret and serve as building blocks for more advanced algorithms like **Random Forests** and **Gradient Boosting Machines.**

Naives Bayes (NB)

NV is a supervised ML algorithm that works with independent assumptions (Bayes Theorem). The theory is based on an assumption that the presence of a particular feature in a class of dataset is unrelated to any other features within the datasets. The fundamental of this approach is rooted on the assumption that the classification of input data is conditionally independent of their features. This allows the algorithm to make predictions quickly and accurately. (Ray, 2025). The Naïve Bayes algorithm is typically known for handling high dimensional data, and feature recognition such as URL and text analysis making it an effective tool for classifying tasks and detecting phishing websites. It also offers a simplistic approach making its fast and efficient.

Random Forest

RF is a combination of several decision trees independently trained on select datasets to enhance prediction and accuracy. Just like decision tree, it is also used for both classification and regression tasks. The RF algorithm is however more effective in detecting phishing websites than a single decision tree due to its ability to combine several decision trees to analyze various data features and extracts from websites information such as length of URL, number of sub domains and the presence of special attributes. The result of its analysis typically reveals whether a website is phishing or legitimate (Wang et al., 2021).

Logistics Regression

LR an ML algorithm primarily used for classification of binary tasks. This makes it a powerful too in differentiating between two classes such as phishing or legitimate websites. LR can predicts the possibility of an input belonging to a particular class using the logistic function with output value between 0 and 1. To detect phishing websites, LR works by analyzing various features of a website, such as URL structure, length, domain information such as its age, and page content. During training, the algorithm learns the relationship between these features and the target class (phishing or legitimate) by optimizing a loss function leveraging techniques such as gradient descent.

Neutral Networks

NN multilayer perceptron is a powerful ML algorithm that is trained to replicate the behavior of a human brain. The neutral

network structure consists of interconnected nodes that are trained to process input data to get the desired outcome. NN is effective in solving complex tasks such as image recognition, modeling complex patterns and analyzing features of webpages easily. The NN algorithms works by analyzing various features and attributes of a website, such as URL structure, domain information, and content, to determine whether it is legitimate or phishing. The algorithm is developed to extract features from webpages such URL length, presence of suspicious characters, domain age, and behavioral patterns.

Recent research has focused on improving neural network-based phishing detection by using deep learning methods like Convolutional Neural Networks (CNNs) for feature extraction from raw URLs or website content. According to a 2021 study by Li et al., NN achieved high rates of accuracy in detecting phishing websites when used to examine URL and content features of webpages (Li et al., 2021). Further research has revealed that the use of hybrid models involving the integration of NN with other algorithms, such as decision trees or SVMs, have been proposed to enhance predictability and performance. Kumar et al. (2022) introduced a hybrid approach combining neural networks with ensemble methods for improved phishing detection. This research will conduct a comparative analysis of the six supervised ML algorithms mentioned. The experimental analysis will be carried out using datasets from University of California Irvine (UCI) ML repository having 30 features and 11055 instances.

Unsupervised and Semi-supervised learning

Unsupervised ML models are considered unsupervised when trained on unlabeled dataset of URLs. Thus, enabling the model to identify patterns and unusual behaviors in the data that could reveal the presence of phishing websites. On the other hand, semi-supervised learning methods combines elements of both supervised and unsupervised learning by training the ML model on a small, labeled dataset of URLs from phishing and legitimate websites. It also learns from an unlabeled dataset to identify new patterns and irregularities in the data. The dataset feeds the ML models with the necessary information about the data to predict future occurrences.

Deep and Ensemble learning

Gresele (2023) noted that Deep learning methods, such as convolutional neural networks (CNNs) or recurrent neural networks (RNNs), can detect phishing URLs by learning features directly from raw data, such as website screenshots or network traffic logs. Ensemble learning method combines multiple machine learning models to improve performance. The Ensemble method is an effective approach to detecting phishing websites because it combines different types of models with varying strengths and weaknesses. Each method has its strengths and weaknesses, and it is necessary to experiment with multiple methods to identify the most effective approach for detecting phishing websites.

Statement of Purpose

As previously mentioned, series of attempt have gone into identifying features of phishing websites that could form basis for which meaningful classification of malicious websites could be based. The effectiveness of these attempts has been limited by incompleteness of the list of features used and misuse of performance metrics. The choice of performance metrics to use in a classification task should depend on the kind of analysis being carried out. For example, feature selection analysis requires a performance metrics that has a balanced response to True (Positive and Negative) and False (Positive and Negative). This inspired the choice of Matthew's Correlation Coefficient (MCC) as the primary evaluation method in our investigation. This research work will address the problem of phishing websites which are a major cyber threat. In addition, this research work will also try to close the gap by addressing existing issues on previous research works on the area, specifically attributed to performance issues of ML algorithms.

Significance of the Study

Despite several attempts to curb phishing website attacks, it remains one of the most prevalent cybersecurity threats. As previously mentioned, traditional phishing detection methods such as DNS filtering, blacklist, and whitelist methods, and user awareness training are not reliable in curbing the problem with phishing websites because of the rapid evolution of phishing techniques. The adaptive and innovative methods used by ML techniques have shown significant improvement in tackling phishing websites when compared to traditional approaches. Is it noteworthy that ML techniques can detect phishing in real time. This helps organizations stay ahead of the problem and mitigate potential risks even before they occur. As previously mentioned, ML algorithms such as SVM, NN, DT and many more highlighted in this study can improve their accuracy in detecting phishing by training using datasets. This reduces the risk of false positives and false negatives.

Deep Learning and Ensemble Learning techniques can further enhance the detection of phishing websites by analyzing their visual and text features, making them more effective in solving the threat associated with phishing. A major upside in using ML algorithms to detect phishing websites is their ability to recognize patterns based on prior experiences and ability to analyze large volumes of data making them effective in detecting zero-day threats and identifying new phishing methods. As previously stated, millions of dollars are lost annually as a result of phishing attacks. These attacks impact both developed and developing economies that rely on digital services for daily transactions. Leveraging ML algorithms to detect phishing websites is a breakthrough, as it helps safeguard users and customers and restore trust in businesses as they interact with them for required services.

Assumptions & Limitations of the Study

The following are assumptions that could lead in inaccuracy in the results for phishing detection using ML models:

- **Reliability of Datasets-** The reliability of datasets is not always accurate. It is based on the assumption that datasets used to train ML models are representative of real-life scenarios from phishing and legitimate websites which isn't exactly the case.
- **Generalization od ML Models:** This is based on the assumption that trained ML models will perform to optimum expectation when applied to other datasets outside the phishing and legitimate websites
- **Consistency of Evaluation Metrics:** It is assumed that select performance matrix such as Accuracy, Receiver Operating Characteristic (ROC), Curve Performance Metrics, MCC, all effectively measure the model's capacity to detect phishing websites.

- **Potential False Positives and Negatives:** There is a risk of legitimate websites being falsely classified as phishing (false positives) and phishing sites being misclassified as legitimate (false negatives).
- **Bias and Incompleteness of Datasets:** Bias could occur if the trained datasets fed into ML models do not contain all possible phishing techniques. New phishing techniques may emerge that the ML model is yet to encounter.
- **Computational Complexity:** Some machine learning models, like deep learning are difficult and impractical for real-time phishing detection in resource-constrained environments.

This rest of this research paper offers insights into six algorithms used for the comparative analysis, a review of the latest research on phishing attacks and outlines the methodology employed in this study. The experimental results of our comparative study are presented and discussed and concludes the paper by summarizing the key findings and proposing avenues for future research.

Chapter 2

Overview

The risk of phishing websites has grown in recent years due to heavy reliance on digital technologies and is considered one of the most prevalent cybercrimes. The potency of web phishing attacks continues to cost millions of dollars in losses with tremendous negative impact on web users. Phishing websites are bogus sites where an attacker attracts unsuspicious victims to a spoofed website with the appearance of a legitimate one. According to Gillis (2024), when victims access a phishing site, the attacker attempts to convince them to disclose their private information, such as usernames, passwords, and credit cards, resulting in theft of personal information.

Phishing websites continue to target web users, online businesses, and government platforms to steal sensitive information. Therefore, identifying these phishing attacks on time is critical to safeguarding users from associated risks. However, detecting a phishing website is challenging due to the many innovative methods phishing attackers use to deceive web users. The success of phishing website detection techniques mainly depends on recognizing phishing websites accurately and within a prompt timeframe.

Many conventional techniques such as DNS filtering, blacklisting and whitelisting databases have been suggested to detect phishing websites. Most of these conventional techniques are not reliable and struggle with detecting whether a website is phishing or legitimate. This has led to many new phishing websites wrongly classified as legitimate websites and vice versa. Conventional techniques are also inefficient with zero-day phishing attacks as threat actors continue to evolve and new phishing websites are launched quickly.

Literature Search Strategy

The literature review for this research piece was sourced from prominent journals, articles, and secondary data published between 2020 and 2025. Some of the sources consulted include Research Gate, BAU Online Library, Wiley Online Library, IEEE, Google Scholar, Forbes Articles, and Reports from the Anti-Phishing Working Group. However, some fundamental research work from earlier years highlighted the evolution of phishing website attacks and how ML algorithms have progressed over time to tackle the phishing problem. The core focus of the literature review is phishing website detection using ML algorithms and related terms. As previously mentioned, the study will compare six ML algorithms (Support Vector Machine (SVM), Decision Tree (DT), Naïve Bayes (NB), Random Forest (RF), Neutral Networks (NN), and Logistics Regression (LR)) to determine their accuracy in detecting phishing websites. The entire content for this research was derived from secondary data sources. This section of work throws insights into collective efforts and research conducted by different authors to curb the problem of phishing websites and what the study aims to achieve:

1. The study focuses mainly on a comparative analysis of phishing websites using ML algorithms. To ensure the analysis was effective, the study delved into different ML algorithms, including (SVM, DT, NB, RF, LR, NN) and how effective they are when compared to traditional approaches such as DNS filtering and the blacklist and whitelist approaches. Basit et al. (2020) stated that the accuracy of these traditional approaches was low, and they could only recognize 20% of phishing attacks. Their study revealed that ML techniques give better outcomes with higher accuracy for phishing website detection, thus reducing false positives and negatives. This research analyzed a phishing website dataset with six supervised algorithms using a feature selection investigation on the most promising of the six algorithms, primarily using the filter method and compared it with the outcome of the wrapper method.
2. The comparison of the various ML algorithms will be classified based on four categoric features including Address Bar Based Features, Abnormal Based Features, Hypertext Mark Language (HTML) and JavaScript based Features and Domain Based Features. The research will use dataset from University of California Irvine (UCI) ML repository having 30 features and 11055 instances dataset.
3. In addition to Accuracy and ROC *(Receiver Operating Characteristic) Curve performance metrics, the research will consider MCC (Matthews Correlation Coefficient).*
4. The research experiment will be carried out using the filter feature and wrapper selection methods on Weka platform 3.8.6 on MacBook Air.
5. The research framework is supported with relevant data to interpret the research accurately.
6. Evidence and accuracy of data measurement and reporting is provided clearly and concisely.
7. The presented data support the conclusions.

During this research, at least a total of 35 articles were reviewed. It is noteworthy that some studies employed multiple techniques for phishing detection, resulting in their inclusion under multiple categories. Of the 35 articles, 29 utilized ML approaches for detecting phishing attacks. Considering these numbers, approximately 71.25% of the research conducted in this field

focused on utilizing ML algorithms, the highest among the six mentioned techniques. Among the machine learning approaches, Deep Learning was the most employed, with 26 articles (66.25%) utilizing this technique. The articles revealed ensemble learning as the most recent hybrid approach currently being explored to enhance accuracy in phishing websites detection.

Conceptual Framework

The detection of phishing websites using ML draws from various fundamental theories primarily underpinned by three theories: Computational Learning Theory (COLT), Information Security Theory, and Decision Theory. These theories provide the basis for understanding how ML models analyze patterns and features within datasets to differentiate phishing and legitimate websites.

Computational Theory uses mathematical methods to train ML algorithms from datasets. Ultimately, the theory of CoLT works by recognizing the performance of ML models including their time complexity and ability to adapt and deploy easily. **CoLT** is a core part of ML applications in phishing detection. First developed in the late 20th century, CoLT uses mathematical frameworks to quantify learning tasks and algorithms to make accurate predictions. Further research by Valiant (1984) introduced the Probably Approximately Correct (PAC) learning framework, which provides the method for evaluating the efficiency of learning algorithms.

Information security encompasses the collaborative efforts by people and organizations to protect information. This is achieved by putting suitable controls to protect information from threats actors. The control types vary for every organization depending on what the information is used for According to Horn et al. (2016) Information security focusses on what protection is afforded to information and how businesses can leverage information to support their business goals.

Information theory is based on the assumptions that information security depends on a complete information classification assessment. This identifies what information is owned by the organization and therefore what information needs to be protected. Information theory helps to categorize information, identifies and segregates which ones are more important than others. That way, organizations can categorize what information requires utmost protection, making it easier to decide quickly the types of controls measures that are most suitable to safeguard user information. Overall information theory explains how information security can be used to identify suspicious patterns on the web. In cybersecurity applications, information security is useful for detecting malicious techniques in phishing websites.

Decision Theory helps to enhance accuracy in binary classification tasks such as distinguishing phishing from legitimate websites and reducing the occurrence of false positives and negatives. Decision Theory is pivotal for ensuring ML models achieve high accuracy in detecting phishing websites by learning the model's behavior based on past experiences and analyzing input data to make informed and effective decisions. According to Berger (1985), when ML models are trained to recognize patterns learnt from past experiences and observed data, they make more accurate decisions. High accuracy in binary classifications translates to less errors or bias that emanate from false positives and false negatives. In detecting phishing websites, ML models can classify phishing websites from legitimate ones with high accuracy based on prior information gathered.

A combination of the three theories above mentioned theories are fundamental in phishing websites detection because they all leverage pattern recognition and data classification to make smarter decisions and enhance effectiveness of ML algorithms. These theories provide a comprehensive foundation for phishing detection using machine learning.

Rationale for Literature Review on Theory Application

Despite several collaborative efforts by researchers and cybersecurity experts to tackle the problem of phishing attacks, it remains one of the most prevalent cybersecurity threats. As stated in Chapter 1, phishing attacks have resulted in millions of losses in dollars globally. If left unaddressed, the risk of phishing could contribute significantly to a global economic downfall. In a bid to find a solution to the problem of phishing, Abdelhamid et al. (2019) developed a method called Enhanced Dynamic Rule Induction (EDRI) or simply rule induction. The method was developed to improve the accuracy of ML models in detecting phishing websites by introducing a set of rules and feature selection from datasets trained to classify phishing websites from legitimate ones. The approach generates a set of rules by extracting decision rules from datasets after identifying patterns and relationships within them. The rules are continuously refined as new datasets are introduced into the model. When detecting phishing websites, these rules help to distinguish legitimate from phishing websites based on specific features in the datasets. Their study revealed that their rule induction method achieved an impressive 93.5% accuracy, reduced false positives and improved overall detection reliability and ultimately outperformed many traditional ML approaches.

The study also deployed the feature selection approaches to enhance ML model performance. The feature selection technique works by identifying and keeping the most prominent features within the datasets that contribute the most to models performance while removing irrelevant features. The process of reducing the number of features within datasets to only the relevant or most prominent ones enhances the performance of ML models without compromising accuracy. This research will use the filter and wrapper feature selection approaches to rank the features in the datasets sourced with UCL ML repository having 30 features and 11055 instances. The filter and wrapper feature selection approaches will be used to determine the most prominent features within the datasets. Each feature in the datasets will be ranked by assigning a score to determine the features that contribute the most to the performance of the six supervised ML algorithms being analyzed. Zhang et al. (2021) conducted a comparative analysis of DTs and deep learning models to evaluate their performance in detecting phishing websites. Their study provided an in-depth analysis of how DTs and deep learning models handle phishing detection and the difference in these approaches when compared to conventional approaches like DNS testing and blacklisting. The key outcome of their experimental analysis revealed that a combination of the principle of CoLT and Decision Theory significantly increases phishing detection rates. Their study highlighted that integrating CoLT and Decision Theory helps ML models in making smarter decisions, improves their overall accuracy and precision to distinguish between legitimate and phishing websites and ultimately, making them outperform

traditional methods, particularly in reducing false positives and improving overall detection reliability. Their discovery increased the accuracy of ML models and helped reduce some of the recurrent problems in phishing detection, such as the fast-paced evolving attack strategies and manipulation of website characteristics.

A breakthrough in phishing detection is URL analysis which typically is achieved through blacklisting. Blacklisting is one of the most popular conventional approaches used in detecting phishing. According to Zhang et al. (2020), URL analysis relies on databases of known phishing sites, such as those compiled by Google Safe Browsing and Phish Tank. The principle of Information security leverages URL and content analysis to mitigate the risk associated with phishing attacks. Despite their success in phishing detection, further research conducted by Verma & Das (2018) reveals that this conventional approach struggles with zero-day phishing attacks because malicious URLs are created very quickly and difficult to identify instantly.

To address this limitation, ML algorithms are trained to analyze and recognized URL features such as domain name, address bar, subdomain, port number and others. Their ability to recognize unique URL features allows the models to easily spot potential phishing sites by identifying abnormal features that do not align with the original pattern they were previously trained to recognize. Information Security theory uses content-based detection methods to analyze the structure and elements of a website, including URL, and JavaScript. Mohammad et al., (2014) in their research stated that these techniques are effective in binary classification due to their ability to assess the similarities between phishing and legitimate websites by examining key attributes and domain inconsistencies. Another research conducted by Marchal et al. (2017) explained that advanced content-based methods use ML models to classify phishing and legitimate sites based on extracted features, thereby improving detection rates.

Another study conducted Fadaei et al. (2020) stated that deep learning approaches such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) can recognize complex patterns in web content and URLs making them even more effective in detecting phishing websites.

Background of Related Work

Several attempts have gone into identifying innovative ways to tackle the problem of phishing websites. The effectiveness of these attempts are sometimes limited by either the presence of redundant features in datasets or the misuse of performance metrics. This brings us to the importance of selecting the right choice of performance metrics for binary classification task and how it affects the performance of ML models. For example, feature selection approaches require a performance metrics with a balanced response to true positive and negatives and false positive and negatives. It is against this backdrop that this research leverages the filter and wrapper feature selection approaches to determine the performance of six ML models and identifies Matthews Correlation Coefficient (MCC) as a reliable performance metric in achieving a balanced response when dealing with uneven datasets. Over the last decade, ML models has emerged an effective tool in phishing websites detection because they are trained to analyze a diverse range of features and patterns, allowing them to quickly recognize irregular patterns within websites. As a result, the instances of false positives and negatives leading to legitimate websites being classified as phishing or vice versa are greatly reduced. Many authors have explored the detection of phishing websites. However, only a few have conducted a systematic literature review, as described below.

Kunju et al. (2020) conducted a survey to detect phishing attacks using ML algorithms. The study reviewed several ML methods used for approaches used to detect phishing websites. The study focused on how ML algorithms like SVM, RF and NN can detect phishing websites by analyzing features such as URL, content analysis and pattern recognition. Despite the recorded success of these algorithms in detecting phishing websites, the study highlighted some limitations with ML algorithms which are often related to reliance on trained static datasets which sometimes leads to false positives and negatives. The study recommended combining ML approaches and rule-based techniques such as the EDRI for more effective results in detecting phishing websites. Other limitations of this survey include the reliance on only 14 studies ranging from the period 2007 and 2019. This does not represent a full spectrum of studies conducted for phishing detection using ML algorithms. Their research did not take into consideration deep learning approaches which are proven to improve the accuracy of phishing website detection. Instead, it focused primarily on theoretical aspects and lacked an in-depth evaluation of the practical deployment challenges faced by ML models in real-world scenarios.

Hassan (2020) analyzed a dataset with 30 features and 2456 instances using ML and feature selection methods. The author experimented with multiple ML algorithms, including SVM, DT (J48), and NB. His results revealed DT as the best performing algorithm among the three algorithms analyzed recording an impressive test accuracy of 95.40%. The results also demonstrated that 14 of the 30 selected features contributed the most to the model's performance in phishing detection. However, this analysis is based on just 2456 instances. To address this limitation, this research wills investigation with a dataset comprising 11055 instances. When deploying the feature selection approach, the author did not monitor how performance increased as the features were removed from the datasets. This research will apply the filter and wrapper feature selection approaches across six algorithms and monitor their performance to check if they improved as redundant features are removed from the datasets.

Basit et al. (2020) reported a survey on AI based phishing detection techniques. Their research used statistical phishing reports to analyze trends and associated risks of phishing attempts and classified anti-phishing evaluations into four categories: Machine Learning, Hybrid Learning, Scenario-based, and Deep Learning. Results of their research demonstrated that ML based approaches produce the best results compared to other conventional approaches. However, the survey focused only on the theoretical aspects of AI but did not extensively evaluate the practical challenges experienced in deploying AI-based phishing detection systems in real-world internet traffic. Given the rapid evolution of phishing tactics, the survey may not fully address emerging threats and the adaptability of AI models to new phishing attacks.

According to Hannousse and Yabiouche (2020), introducing a general scheme is an effective approach to analyze trained datasets for phishing websites detection. Their experimental

analysis highlighted the importance of feature selection in binary classification task such as distinguishing phishing from legitimate websites. The datasets used for their experiment had 87 features and 11,430 URLs. The results from their experimental analysis revealed that when RF is combined with other hybrid models, it achieved a high accuracy of 96.83% in phishing detection. The findings from their research serves as an incredible resource in phishing website detection though it has some limitations. For instance, the constant need to manually update the datasets is a constraint. A dataset assembled in 2020 may not adequately reflect today's phishing ecosystem. This gap potentially limits the accuracy and precision of models trained exclusively on specific datasets.

Abuzuraiq, Alkasssbeh, and Almseidin (2020) conducted a comparison of different AI methods in phishing website detection. Their research explored the benefits of feature selection on trained datasets and how it enhances model performance. They examined the effectiveness of three ML algorithms including (DT/J48, RF, and NN) in detecting phishing websites. They experimented with a balanced datasets with 5,000 legitimate webpages drawn from Alexa and Common Crawl and 5,000 phishing webpages taken from phishing tanks and open tanks, with 48 features from each URL. Using the feature selection approach, they identified 20 prominent features from the 48 features of the URLs with the RF algorithm demonstrating a high accuracy rate of 98.11%. The study, however, has its limitations. For instance, there were an equal number of legitimate and phishing URLs, which does not reflect real-life scenarios. In real-world scenarios, phishing websites are usually less in number than legitimate websites. Hence, this inconsistency may affect the model's performance when applied to real-time situations. This research paper will analyze 30 features from the UCI datasets with 11055 instances using the filter and wrapper selection approaches to determine which features contribute the most to the performance of the six ML algorithms. The use of both the filter and wrapper feature selection approach is to analyze the datasets using different approaches, making the outcome more accurate and the research more robust.

A comparative analysis of different ML algorithms was also conducted by Khan, Khan, and Hussain (2021) across multiple datasets to evaluate their performance in detecting phishing websites. Their study revealed RF and NN as the best performing algorithms in phishing website detection, with an accuracy of over 97%. Though their study offers valuable insights, it has several limitations. The study was conducted primarily with datasets from a single source from the UCL repository. Their reliance on datasets from a single data source may not capture a broad spectrum of phishing techniques, which could lead to poor performance outcomes. The research also did not explore feature selection methods but focused only on evaluating algorithmic performance in controlled environments. Though this research piece also focuses on a single source of dataset from UCI, it delves into the significance of feature selection (filter and wrapper methods) in the performance of six supervised ML algorithms. The results of these analyses will be revealed in Chapter 4.

Divakaran and Oest, (2022) conducted a comparative analysis on the performance of ML algorithms and Deep Learning techniques in detecting phishing websites. Their research highlighted that accuracy of these models is achieved by combining different data types to achieve maximum efficiency. Their work also brought to the fore the pros and cons of these approaches and presented the various ways to employ them in enhancing phishing website detection. Their study revealed that models that rely solely on URL features may be more prone to phishing attacks, as threat actors can craft benign URLs very quickly. Though it is true that analyzing URL content can enhance detection accuracy, this approach is complex in computing due to the time it takes to extract, and process features from webpage content which could lead to delays, impacting performance in real-time phishing detection.

As phishers continue to advance in their approach to bypass detection systems, it is important for researchers and cybersecurity experts to continuously improve on existing detection techniques to strengthen their effectiveness in tackling the risk of phishing. This study reveals that the use of ML algorithms demonstrates improved accuracy and precision in detecting phishing websites when compared to conventional methods. Ongoing research continues to refine these models with a focus on automated feature selection, classifier performance, and development of diverse and dynamic datasets to combat the rapid evolving phishing tactics.

Synthesis of Literature Findings

Phishing website detection using supervised ML algorithms has been studied extensively. Various approaches, such as feature selection, classification techniques and pattern recognition are being explored to enhance accuracy in detecting phishing attacks. Existing literature by many authors, as previously captured highlights the effectiveness of supervised ML algorithms & deep learning approaches. As previously mentioned, this research will analyze the performance of six supervised ML algorithms to determine the most effective in detecting phishing websites.

In recent times, hybrid models like deep learning and ensemble learning have enhanced predictive performance and accuracy in phishing detection. Despite all the collaborate efforts, there are still gaps in the literature concerning the comparative effectiveness of different ML algorithms in divergent real-world scenarios. The studies revealed that the reliance on specific datasets or recognition of limited features by various ML algorithms still limits accuracy and makes generalization across different phishing attacks difficult. Due to the rapid evolution of phishing techniques, there is a need for continuous model updates to detect phishing attacks. However, existing research often lacks an analysis of model adaptability and robustness against phishing attacks. This study will focus on a comprehensive comparative analysis of six ML algorithms (SVM, decision trees, naive bayes, logistics regression, neural networks and random forest) to bridge this literature gap. The experiment will evaluate their performance on 30 datasets from UCI repository and assess their accuracy and predictive performance across different phishing tactics.

The comparison of the various ML algorithms will be classified based on four categoric features including Address Bar Based Features, Abnormal Based Features, HTML and JavaScript based Features and Domain Based Features. The research will use dataset from UCI machine learning repository having 30 features and11055 instances dataset. This research provides deeper insights into the trade-offs between model complexity and detection effectiveness. Additionally, it investigates the adaptability of these

models in detecting emerging phishing threats, contributing to the development of more resilient phishing detection ecosystems.

This study adds to the existing pool of knowledge in phishing detection by further exploring a more holistic understanding of ML-based phishing detection techniques. This work will help cybersecurity professionals in selecting the most effective ML algorithms to avert phishing attacks in real-world scenarios especially as phishing techniques continue to evolve rapidly.

Summary

Phishing detection using ML algorithms revolves around several major themes. One such is the classification of phishing websites based on feature extraction techniques. Existing studies have explored various feature categories, including URL-based, content-based, and behavioral features, to differentiate phishing sites from legitimate ones. Supervised ML approaches, such as SVM, decision trees (J48), random forests, neural networks, logistics regression, and naive bayes have demonstrated significant accuracy in detecting phishing attacks. The literature reveals that fewer features within datasets achieves higher accuracy. The literature on touches on how hybrid models such as deep learning and ensemble learning models have also contributed to enhancing accuracy in phishing detection.

A comparative analysis of different ML algorithms is crucial in determining the best models to deploy in real-time phishing detection. However, a recurring challenge identified in the literature is the adaptability of ML models to evolving phishing techniques. As previously mentioned, existing studies recognize the role of traditional models such as DNS filtering, blacklisting, and whitelist in phishing website detection. They however highlighted their limitations in binary classification tasks such as differentiating between phishing and legitimate websites.

Despite extensive research, there is a need to deepen the comparative analysis between different ML algorithms, especially with regards to computational efficiency, interpretability, predictive analysis and real-time accuracy detecting phishing attacks. There are existing arguments that while hybrid models like deep learning models achieve high accuracy, they require complex computational resources, making them expensive and more difficult to deploy in environments with limited resources. As phishing techniques continue to evolve, there is need for robust and adaptive ML solutions that can detect threats with minimal human intervention while maintaining high precision and accuracy rates.

Chapter 3

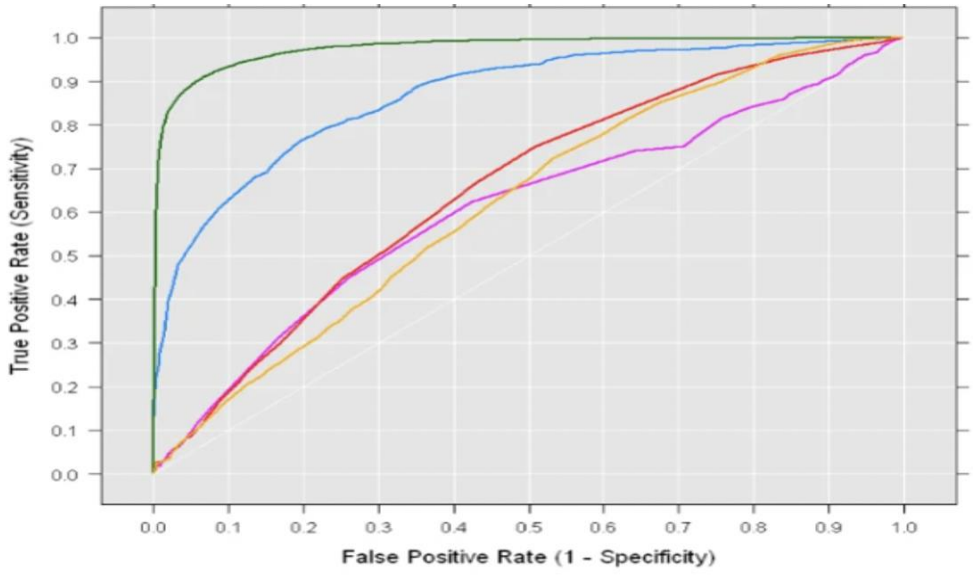
Introduction

So much work has been done in the attempt to identify phishing websites though the results may not be as plausible as expected. Hence, the use of ML algorithms might lead to better classification and choice of evaluation method that supports the kind of analysis that will make a huge difference. This research will focus on a comparative analysis of six supervised ML algorithms especially the ones not used in previous work. It is interesting to note some new insights were discovered.

The use of Accuracy and maybe ROC as performance metrics have become so common that it seen more as the gold standard for binary classification evaluation. The ROC curve as illustrated in Figure 1 shows true positive rate (also called sensitivity or recall) on the *y* axis and false positive rate on the *x* axis, and the ROC area under curve (AUC) ranges from 0- 1 with 0 representing the worst result and 1 the best result). Though ROC matrix is commonly used in binary classifications such as detecting phishing from legitimate websites, it has several limitations. For instance, the score is generated to include predictions with insufficient recall rates. It also doesn't mention anything about *positive predictive value* (also known as *precision*) or negative predictive value (NPV) obtained by the classifier. Hence results from ROC can be overly optimistic and lead to errors and false predictions. (Chicco & Jurman, 2023).

Figure 1

Graphical representation of the ROC



The review of evaluation methods based on this research points us in a different direction especially for binary classification. The MCC was first introduced by B.W. Matthews to assess the performance of protein secondary structure prediction in medicine. It considers true and false positives and negatives and is generally regarded as a balanced measure which can be used even if the classes are of very different sizes". The MCC evaluation matrix will only generate a high-performance score in if the classifier scored a high value across the four basic rates of the confusion matrix: sensitivity, specificity, precision, and negative predictive value. A high MCC (for example, MCC = 0.9), always corresponds to a high ROC AUC, and not vice versa.

The MCC can be seen as a discretization of the Pearson correlation for binary variables. If x and y are binary, using some algebra to represent True Positive as TP, True Negative as TN, False Positive as FP, and False Negative as FN, we can represent

$$MCC(x,y) = \frac{TP*TN-FP*FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$$

Evaluation of MCC figure against Accuracy and ROC as we would see soon clearly shows we made a good decision. This has inspired the choice of MMC as the primary evaluation method in our experiment.

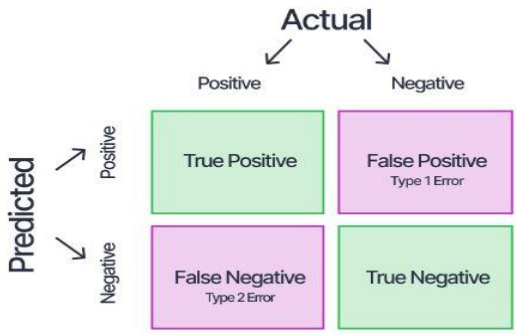
As previously stated, the MCC score captures all the four categories of a confusion matrix. To enhance clarity, a confusion matrix is used in binary classifications tasks to demonstrate the performance of ML algorithms by comparing the model's performamce to the desired results in a table. The matrix breakdowns the results into four categories: true predictions (true positives and true negatives) and false predictions (false positives and false negatives). For instance, in medicine, the confusion matrix can be used to classify benign and malignant cancers by revealing the correct counts of positives (correctly identified malignant tumors), true negatives (correctly identified benign tumors), false positives (benign tumors incorrectly classified as malignant), and false negatives (malignant tumors incorrectly classified as benign). In cybersecurity and ML, the confusion matrix plays a key role in as it helps to detect phishing from legitimate websites. The confusion matrix helps to understudy the performance of ML models by revealing correct and incorrect predictions in a simplified way. The matrix displays the number of instances produced by the model on the test data.

- True Positive (TP): Correct prediction of a positive outcome
- True Negative (TN): Correct prediction of a negative outcome
- False Positive (FP): Incorrect prediction of a positive outcome (Error Type 1)
- False Negative (FN): Incorrect predicted a negative outcome (Error Type 2).

A tabular representation of the confusion matrix in captured in Figure 2 below

Figure 2

A Confusion Matrix



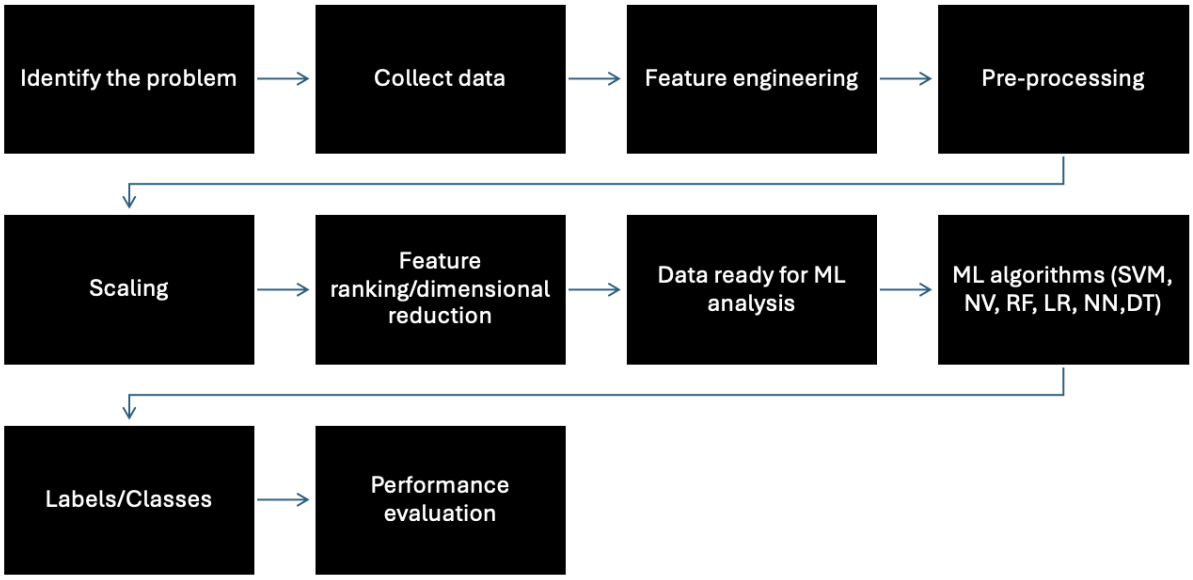
Rationale for the Research Approach

The dataset for this research was gotten from UCI ML repository having 30 features and11055 instances dataset. The use of a large dataset for UCI ML repository is to evaluate their performance across six ML algorithms. This study highlights the features of the datasets and its independent variables and data labels or classes (phishing or legitimate) as its dependent variables. This research aims to also buttress the important of training a diverse range of datasets to optimize performance of ML models. The approach will subject the datasets from UCI to six supervised machine learning algorithms to gain new insight into what algorithm holds the best promise for our dataset, with all the 30 features. With the best performing algorithm, a feature selection investigation (using both the filter and the wrapper methods) to know which of the 30 features could be safely left out of the list without incurring a significant drop in performance. This is done by removing one feature at a time and observing the changes in performance. The purpose of feature selection is to remove redundant features from the dataset that may reduce their performance.

Feature selection includes removal of features that could lead to loss in computational time, noise reduction and irrelevant features to improve an ML algorithm prediction and accuracy. Despite the important of feature selecting in phishing website detection, a challenge associated with the method for phishing websites detection is the need for manual feature selection for enhanced accuracy and precision. As previously mentioned, this study will use the filter and wrapper feature selection approaches highlighted below to identify the most relevant features in the datasets. A challenge experienced during this research is the lack of reliable training datasets. Though, there are many articles and studies on predicting phishing websites using data mining techniques, there is a lack of published & reliable training dataset. This may be because there isn't an agreement in literature on the definitive features that characterize phishing websites. Thus, it is difficult to find datasets that covers all possible features. The result of the experiments will be revealed in Chapter 4. Figure 3 below demonstrates a pictorial representation of the ML pipeline.

Figure 3

The ML pipeline



Methodology

This research employs both qualitative and quantitative methods to address the research questions. The first part of the question was addressed using a qualitative approach to identify six major ML algorithms that demonstrate high accuracy and precision in detecting phishing websites. The second part of the research question was analyzed using a quatitative approach to conduct an experimental analysis on the performance evaluation of the six ML algorithms in detecting websites and how feature selection impacts the performance of the algorithms. To this end, this research will use the filter and wrapper feature selection approaches to rank the features in the datasets and identify the ones that contribute the most to the performance of the six ML algorithms.

Filter Method in Phishing Detection: Filter selection method involves ranking the features of the input datasets to extract the most prominent features by assigning a score to every feature. When assigning the score during the experiment, a statistical measure was applied by the filter feature selection methods on Weka platform 3.8.6 on MacBook Air. The score determines the most relevant features that should be retained and the redundant features to be removed from the datasets. The removal of redundant features is important to avoid slowing down and confusing the algorithm. The filter selection method is usually univariate and take the feature into consideration independently, or with regard to the dependent variable. During the experiment on Weka, the filter selection feature technique was implemented by using Information Gain (IG). IG is a crucial measure used for ranking datasets. it also measures the extent to which the features within a dataset are mixed up. (Taminu J et. al, 2024).

Wrapper Feature Selection Methods: For the purpose of this research, the wrapper selection feature classifier is used to select the most prominent features from input datasets necessary to distinguish phishing from legitimate websites. In wrapper feature selection, the machine classifier is the main ingredient use to extract the most relevant features in the input datasets. This research will focus on a comparative analysis of six supervised ML algorithms (SVM/SMO, RF, DF/J48, NN, NB, LG) to determine the most effective in phishing websites detection.

The experimental analysis will rank the features of datasets using the filter selection approach to highlight the best features that contribute the most to the performance of the models when classifying phishing from legitimate websites. The datasets from UCI ML repository will be uploaded onto Weka 3.8.6 and subjected to four experimental analyses to evaluate the six ML models performances. During the initial experiment, the datasets will be subjected to an (80%) (20%) split with 80% representing the training datasets and 20% test datasets. The percentage split approached was first selected to assist in separating the training datasets from the test datasets. The test will be conducted across all the six ML algorithms to identify the best performing model.

The experiment will be repeated a second time, but this time around, the training datasets will be increased by (10%), making a percentage split of (90%) training datasets and (10%) test data. The increased in the training datasets is to determine if analyzing a larger dataset can lead to improved prediction of the models. Following the percentage split approach, the datasets will again be subjected to a cross validation approaches first with a five-fold. After which they will again be subjected to 10-fold cross validation.

Finally, the experiment will explore the performance of the datasets vs number of features curve. The result of the experimental analysis will address the research's questions- what ML techniques are most effective in identifying phishing websites, and how they perform in the context of cybersecurity threats.

Design

All the previous work considered during this research accept that the 30 data features can be grouped into 4 categories *(Address Bar Based Features, Abnormal Based Features, HTML and JavaScript based Features and Domain Based Features)* as illustrated in Table 1 below.

Table 1

30 Extracted features from Phishing Website dataset UCL ML repository

Feature Category	Feature Names	Number
Address Bar based Features	Using the IP Address, URL-Length, Shortining-Service, having-At-Symbol, double-slash-redirecting, Prefix-Suffix, having-Sub-Domain, SSLfinal-State, Domain-registration-length, Favicon, port, HTPPs Token	12
Abnormal Based Features	Request-URL, URL-of-Anchor, Links-in-tags, SFH, Submitting-to-email, Abnormal-URL	6
HTML and JavaScript based Features	Redirect, on-mouseover, RightClick, popUpWidnow, Iframe	5
Domain based Features	age-of-domain, DNSRecord, web-traffic, Page-Rank, Google-Index, Links-pointing-to-page, Statistical-report	7
Total Features		30

The comparative analysis conducted by Khan, Khan, and Hussain (2021) across multiple datasets to analyze their efficacy in detecting phishing websites revealed that Random Forest and Artificial Neural Networks as the best performing algorithms with over 97% accuracy. Though their study was conducted primarily with datasets from the UCL repository as we have done as well, their research also did not explore feature selection methods but focused only on evaluating algorithmic performance in controlled environments.

Abuzuraiq, Alkasssbeh, and Almseidin (2020) on the other hand conducted a comparison of different AI methods in phishing websites detection. Their study examined the effectiveness of three ML algorithms (Decision Tree/J48, Random Forest, and Multilayer Perceptron) in detecting phishing websites with RF achieving a 98.11% accuracy. Their decision to only evaluate three algorithms lives a gap in research as it pertains to other prominent ML algorithms. They also didn't apply the filter or wrapper-based selection methods. Our experiment will cut across six ML algorithms including predictive analysis using the filter and wrapper feature selection methods.

Data Analysis Plan

As previously stated, the WEKA 3.8.6 data mining tool will be used for the experimental analysis. The datasets gotten from UCL ML repository was pre cleaned and ready for use. The experiment focused on a comparative analysis of six ML algorithms to identify the best performing model for detecting phishing from legitimate websites. The results of the performance evaluation of the six algorithms will cut across seven performance indicators of which four are derived from the confusion matrix. The seven performance indicators include accuracy, F1 score or measure, precision, recall, MCC, ROC, and precision recall curve (PRC). These indicators are critical for this comparative analysis because they provide a quantitative measure of a model's performance. To enhance clarity of the results, the definitions of the key performance matrix are further highlighted below:

- Accuracy: The proportion of total predictions that were correct (both true positives and true negatives).
- Precision The proportion of positive identifications that were correct. It is a measure of the model's accuracy in predicting the positive class.
- Recall: The proportion of actual positives that were identified correctly. It indicates the model's ability to detect all positive instances. (True positives)
- F1- Score: The weighted average of precision and recall, providing a balance between the two metrics. It is particularly useful when the class distribution is imbalanced. (The F1- score is particularly important in this experimental analysis because there is an imbalance in the training and test datasets which is reflected of real time internet traffic where legitimate websites out number phishing websites.
- MCC- The key performance matrix for this research analysis
- ROC- A common performance matrix used to evaluate ML models
- PRC- The precision recall curve is the measure of the area under curve of precision and recall. This can also be referred to as area under curve (AUC).

Following the comparative analysis of the six ML algorithms, the datasets will be subjected to the filter selection feature method to rank the data features by assigning a score to each feature according to their contribution to the models performance. The final experiment will use the wrapper feature selected approach to evaluate the datasets and select the most prominent features. The results from the experiments conducted on Weka will be exported and illustrated using excel spreadsheets and graphs for ease of interpretation. The results from the experiments will demonstrate the most performing ML algorithm for phishing websites detection. It will also highlight the best features within the datasets that contribute the most to accuracy and precision in six ML algorithms.

Chapter 4

Introduction

This research aims to analyze six ML algorithms to determine the best performing algorithm in phishing websites detection. The research also seeks to address two critical questions- What machine learning techniques are most used to identify phishing websites, and how do ML algorithms perform in regard to identifying phishing websites in the context of cybersecurity threats? Among the six machine learning algorithms analyzed, there will be significant variations in their performance for phishing website detection, with one algorithm outperforming the others based on some key performance matrix such as accuracy, MCC, ROC, precision, recall, and F1-score. It is important to note that these six ML models will demonstrate distinct strengths and weaknesses in identifying phishing websites within the context of cybersecurity threats. Hence, this research seeks to address two critical questions- What machine learning techniques are most used to identify phishing websites, and how do ML perform in regard to identifying phishing websites in the context of cybersecurity threats?

In this Chapter, a detailed review of the results from the experimental analysis of the six ML algorithms will be revealed. To conduct a comparative analysis of the six supervised ML algorithms ((SVM/SMO, RF, DF/J48, NN, NB, LG), four experiments were conducted on Weka 3.8.6 data mining platform. The first two experiments were conducted by subjecting the datasets for UCI to percentage split approach while the third and fourth experiments were conducted using a cross-validation approach. Details of the experiments, the results and analysis therefore are highlighted below.

Data Results & Analysis

To have access to multiple supervised machine learning algorithms we ran our experiments on Weka platform 3.8.6 on Macbook Air The initial experiment was conducted across six supervised machine learning algorithms (SVM/SMO, RF, DF/J48, NN, NB, LG), to identify the most promising of all the algorithms used. We used the full dataset for this and below are the evaluation figures for each of the algorithm. The initial experiment with datasets from UCI was ran through WEKA using a random sampling. The datasets were split (80%) (20%) with 80% representing the training datasets and 20% test datasets. The result of the initial experiment across the seven performance indicators/metrics are captured table 2 below:

Table 2

Evaluation of Initial Analysis with 6 Algorithms

Initial Model Performance							
Algorithm	Accuracy	Fscore	MCC	ROC	PRC	Precision	Recall
SVM	93.62	0.936	0.871	0.934	0.908	0.937	0.936
LG	94.53	0.945	0.89	0.987	0.988	0.946	0.945
DT/J48	96.67	0.967	0.933	0.99	0.989	0.967	0.967
RF	97.16	0.971	0.943	0.996	0.996	0.972	0.972
NB	92.85	0.928	0.856	0.0979	0.98	0.929	0.929
NN	94.52	0.945	0.89	0.987	0.988	0.946	0.945

The experiment revealed Random Forest as the best performing algorithm with an accuracy of 97.16% in phishing websites detection and an MCC score of 0.943 and F score of 0.971. The experiment was repeated a second time with 90% training datasets and 10% test datasets. This was done to check if a large dataset will lead to higher performance of the ML models. The result of the second analysis is captured in the table 3 below:

Table 3

Evaluation of Second Analysis with 6 Algorithms

2nd Model Performance							
Algorithm	Accuracy	Fscore	MCC	ROC	PRC	Precision	Recall
SVM	93.76	0.937	0.876	0.936	0.91	0.939	0.938
LG	94.21	0.942	0.885	0.987	0.988	0.943	0.942
DT/J48	95.93	0.959	0.919	0.989	0.986	0.96	0.959
RF	97.23	0.973	0.946	0.997	0.997	0.973	0.973
NB	92.85	0.928	0.858	0.979	0.98	0.93	0.929
NN	96.74	0.967	0.936	0.993	0.993	0.968	0.967

From table 3 we can see that the Random Forest again has the best MCC figure (0.946) and accuracy of 97. 23% and F score of 0.973. This time around, we noticed improved performance again for Random Forest of 97.23% against 97.16% earlier

recorded during the initial experiment which we believe is due to use of larger dataset.

Following the first two experiments, another set of experiments were conducted and analyzed on Weka using the six algorithms, but this time with the cross-validation approach. The first experiment leverages a cross validation 5-fold spit to evaluate the model's performance and generalization ability. Similar to the percentage split approach, the cross-validation helps to identify the best performing model with regards to accuracy, precision (or any metric) being analyzed using the datasets. Using the 5-fold validation approach on Weka, the datasets was divided into five equal sets (as referred to as the folds) for training, testing and validation. The results of the 5-fold cross validation is captured in table below. The results from the 5-fold split again revealed Random Forest as the best performing algorithm with a 97.20% accuracy, an MCC score of 0.943 and an F score of .0972.

Table 4

Evaluation of Initial Analysis with 6 Algorithms Using a Cross Validation Approach

Cross validation 5 Folds							
Initial Model Performance							
Algorithm	Accuracy (%)	Fscore	MCC	ROC	PRC	Precision	Recall
SVM	94.73	0.937	0.873	0.935	0.909	0.937	0.937
LG	93.90	0.939	0.876	0.987	0.987	0.939	0.939
DT/J48	95.81	0.958	0.915	0.982	0.977	0.958	0.958
RF	97.20	0.972	0.943	0.996	0.995	0.972	0.972
NB	92.91	0.929	0.856	0.981	0.982	0.929	0.929
NN	96.53	0.965	0.930	0.993	0.993	0.965	0.965

Again, the datasets were subjected to a cross validation 10-fold split. In this instance, the datasets were split into 10 equal sized sets. Each set is divided into two groups: 90 labeled data for training and 10 labeled data are used for testing. it produces a classifier for each of the six algorithms from 90 labeled data and applies that on the 10-testing data for set 1. It does the same thing for set 2 to 10 and produces 9 more classifiers and averages the performance of the 10 classifiers produced from 10 equal sized (90 training and 10 testing) sets. The result of the cross validation 10-fold split is captured in table 5 below:

Table 5

Evaluation of Initial Analysis with 6 Algorithms Using a Cross Validation Approach

Cross validation 10 Folds							
Initial Model Performance							
Algorithm	Accuracy (%)	Fscore	MCC	ROC	PRC	Precision	Recall
SVM	93.80	0.938	0.874	0.936	0.910	0.938	0.938
LG	93.99	0.940	0.878	0.987	0.987	0.94	0.94
DT/J48	95.88	0.959	0.916	0.984	0.98	0.959	0.959
RF	97.27	0.973	0.945	0.996	0.995	0.973	0.973
NB	92.98	0.93	0.858	0.981	0.982	0.93	0.930
NN	96.86	0.969	0.936	0.995	0.995	0.969	0.969

Using the 10-fold split cross validation approach, Random Forest yet again emerged as the best performing algorithm with a 97.27% accuracy, an MCC score of 0.945 and ROC of 0.996.A key observation from the experiments using both the percentage and cross validation split approaches is that having a large training datasets yields more accurate results.

After the first sets of experiments to evaluate the performance of the six algorithms, the filter feature selection method was then used to generate ranking of the features in the order of how much impact they have on the classification tasks. See Table 6 for the ranks as generated using the filter feature selection method. From the table, it is evident that SSLfinal_State is ranked first which means it's the most important attribute in this classification task while popUpWindow is ranked lowest. With the features ranked, we then starting from the least important feature removed features one at a time to see how classification using RF behaves when we have less than 30 features. Table 6 below contains evaluation of experiment carried out starting from the top 31 features and ended at top 14 features. Figure 1 presents a curve of performance vs no. of features.

Table 6

Ranks Generated with Filter Method

S/N	Rank	Feature Number	Feature Description
1	0.4994828	8	SSLfinal_State
2	0.4773031	14	URL_of_Anchor
3	0.1234319	6	Prefix_Suffix
4	0.1145921	26	web_traffic
5	0.1097354	7	having_Sub_Domain
6	0.0470371	13	Request_URL
7	0.0368013	9	Domain_registeration_length
8	0.0374905	16	SFH
9	0.0470371	15	Links_in_tags
10	0.0119270	28	Google_Index
11	0.0106639	24	age_of_domain
12	0.0080047	27	Page_Rank
13	0.0063791	1	having_IP_Address
14	0.0045681	30	Statistical_report
15	0.0041229	25	DNSRecord
16	0.0033867	3	Shortining_Service
17	0.0043661	29	Links_pointing_to_page
18	0.00267231	18	Abnormal_URL
19	0.0020109	4	having_At_Symbol
20	0.0527867	2	URL_Length
21	0.0418384	20	on_mouseover
22	0.0398539	12	HTTPS_token
23	0.0386076	5	double_slash_redirecting
24	0.0364189	11	port
25	0.0201135	19	Redirect
26	0.018249	17	Submitting_to_email
27	0.0126532	21	RightClick
28	0.0033935	23	Iframe
29	0.0002795	10	Favicon
30	0.0000859	22	popUpWidnow

Table 7

Ranked Features Analysis Evaluation

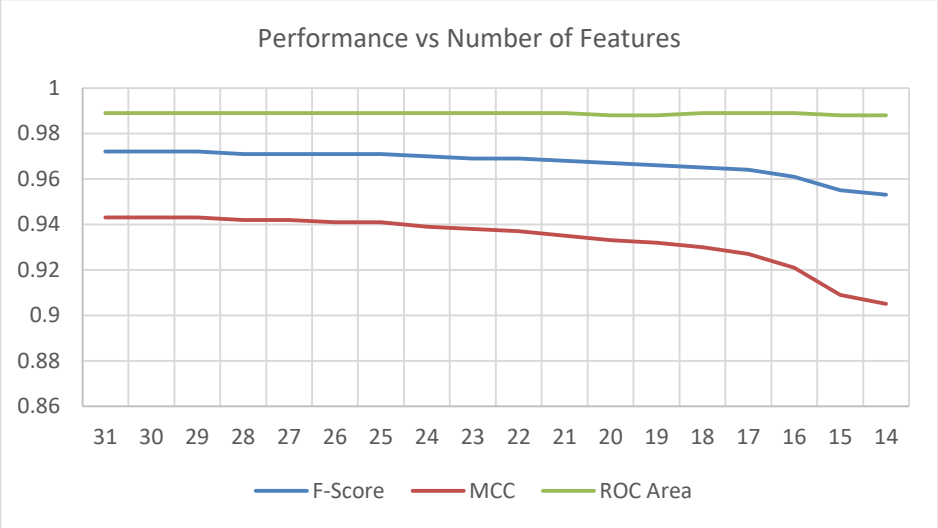
No of Attributes	F-Score	MCC	ROC Area	Accuracy
31	0.972	0.943	0.989	0.971777
30	0.972	0.943	0.989	0.971958
29	0.972	0.943	0.989	0.971777
28	0.971	0.942	0.989	0.971416
27	0.971	0.942	0.989	0.971416
26	0.971	0.941	0.989	0.970692
25	0.971	0.941	0.989	0.970692
24	0.97	0.939	0.989	0.969697
23	0.969	0.938	0.989	0.969516
22	0.969	0.937	0.989	0.968702
21	0.968	0.935	0.989	0.968069
20	0.967	0.933	0.988	0.966893

19	0.966	0.932	0.988	0.96626
18	0.965	0.93	0.989	0.965445
17	0.964	0.927	0.989	0.963998
16	0.961	0.921	0.989	0.961194
15	0.955	0.909	0.988	0.955133
14	0.953	0.905	0.988	0.953053

We lastly used wrapper method feature selection method to evaluate the dataset which yielded the following 21 selected features: 1,2,3,4,6,7,8,10,13,14,15,17,20,21,24,25,26,27,28,29,30. Experiment with these set of features resulted in an MCC evaluation figure of 0.931 which roughly matches use of top 19 features of the ranked filter method. Figure 3 below shows a graphical representation of the performance of the datasets versus the number of features.

Figure 3

Performance vs Number of Features Curve



Summary

Machine learning (ML) algorithms have significantly enhanced detection of phishing websites. Experimental analysis of the six aforementioned supervised ML algorithms revealed that algorithms such as RF, NN and DT have demonstrated high accuracy rates of 95-97%. Their success is due to their ability to recognize patterns, behavioral analysis, feature analysis such as URL structures and webpage content. These algorithms also have track record and strong performance in binary classifications tasks such as distinguishing phishing from legitimate websites. They are also able to train datasets containing both legitimate and malicious URLs by analyzing attributes such as domain age, hyperlink anomalies, and keyword frequency.

Despite recorded success in phishing detection by these algorithms, threat actors continue to deploy innovative ways to alter page elements to evade detection. As previously mentioned, ML algorithms still struggle with detecting new or ero-day phishing attacks, making it easy to miss out detecting attacks that use new tactics before the models are trained. Despite these flaws, ML algorithms remain a powerhouse of cybersecurity contributing significantly to providing solutions to cyber threats. Several ML techniques dominate phishing website detection, each with distinct strengths and limitations in cybersecurity applications.

As revealed in the experimental analysis conducted, RF for instance demonstrated high accuracy of over 97% in detecting phishing websites due to its ability to handle large features. In this instance with 30 features and 11055 instances or attributes, such as URL length, domain name, and DNS records. It performs well with imbalanced datasets and provides feature importance feature rankings, making it a preferred choice for real-time detection systems**.** On the reverse, though SVMs demonstrated accuracy of 94%, they struggle with handling large-scale datasets. In the context of cybersecurity, ML models are faced with challenges like threat evasion hence there is need for frequent training of these models to adapt quickly and counter new or zero-day attacks. While simpler models RF are favored for deployment due to their speed and transparency, deep learning offers superior accuracy at the cost of complexity. Future advancements may focus on explainable AI to improve trust in automated detections and enhance privacy-preserving threat analysis.

Chapter 5

Summary and Conclusion

Introduction

As stated in the previous chapter, threat actors continue to design phishing websites to look like legitimate ones. The aim is to steal sensitive and personal information from internet users such as banking information, passwords or personal data (Guo et al., 2025). ML algorithms can reduce the occurrence of phishing attacks by analyzing various web features such as URL, domain name, domain address, webpage content to detect abnormal patterns. This is achieved by training ML algorithms on datasets containing both phishing and legitimate websites, allowing them to detect phishing websites with high accuracy. The use of ML algorithms to detect phishing attacks play a crucial role in cybersecurity by providing effective mechanisms to detect and prevent evolving phishing threats.

The focus of this research is a performance evaluation of six ML algorithms (SVM, RF, DT, LG, NN and NB to determine their effectiveness in phishing websites detection. The experimental analysis, conducted on Weka 3.8.6 data mining tool leveraged the percentage split and cross validation approaches to analyze the datasets and identify the most prominent features in the datasets. The analysis includes a comparison of their accuracy, MCC, RoC, precision, recall, and PRC which are critical indicators to determine the most effective approach for detecting phishing websites.

Summary of the Results

Following the experimental analysis of the performance of six ML, the following observations were drawn:

- From the performance vs number of features curve captured in Figure 3 of Chapter 4, the ROC Area appears to be not so sensitive to incorrect classification. Accuracy and F-score are better, but MCC shows the best reaction. It is also interesting to see F-score and accuracy curves overlap
- Looking at MCC curve, it is evident that until 25 features the performance seemed not to change as much. We can infer from this that the first 25 features give us the right blend of generalization and performance.
- The ranking of the 30 features from the datasets using the filter selection approach shows that 18 features contribute the most to the performance of the algorithms. It is therefore safe to conclude that 18 features will generate more accurate results than all the 30 features since 11 are mostly redundant features.
- The analysis of the six algorithms using the filter and wrapper feature selection approaches reveals that 25 attributes and larger dataset would yield even better performance.

Interpretation of Findings

Existing literature as quoted previously in Chapter 2 and results from the experimental analysis in Chapter 4 reveal that feature selection contributes significantly to performance of ML algorithms. The findings of this study reinforce the current understanding of phishing websites detection using ML algorithms. As previously mentioned, this research leveraged key research from the past five years (2020-2025) that highlight the effectiveness of different ML models such as RF, DT, SVM in recognizing phishing patterns through features analysis, and pattern recognition. This study confirms these observations by demonstrating that ML models, especially RF, demonstrates high accuracy, recall and precision in detecting phishing websites when compared to others such as DT, SVM, NN, NB and LG. It also showcases MCC as a reliable performance metrics for binary classification tasks. The RF algorithm combines the strengths from multiple DTs to improve its performance. The study adds to the existing pool of reports from notable researchers mentioned in existing literature with similar observations on different ML algorithms and their applications for phishing websites detection.

This findings of this study on phishing websites detection highlights the benefits of using feature selection (filter and wrapper approaches) to test the performance of ML algorithms. From the experimental analysis in Chapter 4, it is evident from the filter feature selection approach that the six ML algorithms can function optimally with 18 features from the 30 features in the datasets. When the results from the filter and wrapper methods were analyzed, it was safe to say that 25 features contributed the most to the performance of the algorithms and the remaining 5 features were completely redundant within the datasets. Findings from the experimental analysis also showed that larger trained datasets yield better results. This study also touches on how hybrid methods like ensemble and deep learning approaches are improving the process of binary classification tasks. As stated previously in chapter 1, deep learning techniques such as CNNs, RNNs and ensemble learning contribute to enhancing phishing website detection by bringing in more advanced and accurate ways to analyze data. It is important to note that unlike ML algorithms, deep learning models can automatically recognize complex patterns in large datasets without the need for manual feature selection. This is especially useful in phishing detection because phishing websites often use subtle tricks that conventional models struggle to detect.

Limitations of the Study

Though this research provides further insights into the effectiveness of ML algorithms for phishing website detection, there are various limitations that affect its reliability, validity, and user trust. One of such is the reliance on pre-existing or pre trained datasets. Pre-trained datasets are limiting because they do not represent the broad spectrum of the evolving nature of phishing attacks. The risk of overfitting is also eminent with pre-trained datasets because they learn the data patterns so much, which potentially leads to poor performance anytime they encounter new phishing patterns not previously trained to recognize. Pre trained datasets also become obsolete quickly. This makes it difficult or almost impossible for them to meet up with the fast-paced and innovative techniques phishers deploy to outsmart phishing detection. (Daniel et al., 2025).

Another challenge with the use of pre trained datasets is the lack of exposure to real life internet traffic. During this research, the experimental analysis for the six ML algorithms was conducted in a controlled environment on Weka (3.8.6) using 30 datasets with 11055 attributes. The results from the experimental analysis revealed over 90% accuracy across the six ML algorithms. However, there is no guarantee that these algorithms will achieve the same level of high accuracy when exposed to real live internet

traffic where such controls on Weka do not apply. This limitation has a ripple effect that could lead to a lack of trust in the findings. Though the six ML algorithms all achieved over 90% high accuracy in experimental settings on Weka, their effectiveness when exposed to real time phishing is yet to be determined.

How the results of the experimental analysis on Weka are interpreted also poses a challenge. While the experimental analysis of the six algorithms conducted on Weka offers some transparency, the platform did not provide an explanation of how the results were achieved. This makes it difficult to understand how the decisions were made. Finally, adversarial machine learning is a limitation to this research. The time allotted for this research was insufficient time to scan the datasets and remove any rogue inputs or data that may have been deliberately manipulated or inserted to introduce bias into the datasets or ML models.

Despite these limitations, the study contributes to the understanding of ML applications in phishing detection. To ensure the battle against phishing attacks is won, future research should address these constraints by incorporating dynamic datasets instead of static datasets. There is a need to also incorporate real-world testing, and standard evaluation protocols to enhance the performance of ML algorithms.

Recommendations

Given the strengths and limitations of the current study, the following recommendations are crucial to improve future research in phishing detection. For instance, this research used sole sourced pre trained datasets from UCI ML repository. This creates a limitation because when there is a new phishing attack, ML algorithms can easily miss on detecting the attack if the datasets have not been trained to recognize the pattern. This necessitates further research to capture a more recent and diverse range of datasets that can quickly identify evolving phishing. There is also a dire need to make balance datasets readily available and published to prevent bias toward specific phishing techniques like URL-based and content-based attacks. (Kulkarni et al., 2024). This research relied solely on static datasets from UCI, which creates limitations to adaptability. Future research should consider the use of dynamic datasets and web crawlers to gather real-time phishing samples.

Though this research touched on hybrid approaches such as deep learning and assemble learning, further deep delve is required. The rapid evolving phishing ecosystem necessitates the need to explore other ways to tackle phishing problems such as combining the strengths of different algorithms to enhance their performance. Further research should explore how semi-supervised or self-supervised learning handle limited labeled data.

Conclusions

Phishing websites continue to pose a significant threat to cybersecurity, exploiting unsuspecting users and compromising sensitive data. Machine learning has emerged as a powerful tool in detecting these malicious websites by analyzing various features such as URL structures, content-based attributes, and behavioral patterns. Through a comparative analysis of different six ML, it is evident that each model has unique strengths in terms of accuracy, MCC, RoC, precision, recall, and computational efficiency.

The ability for ML and deep learning models to recognize abnormal patterns by analyzing web features makes them powerful tools for detecting phishing attacks. Ensemble learning achieves high accuracy by combining the strengths or predictive abilities of multiple ML models to improve phishing detection. This approach is effective in phishing detection because it reduces, and the weaknesses found in individual models. Ensemble techniques also reduce the occurrence of overfitting in trained datasets and improves the model's ability to adapt to new or evolving phishing tactics. ML, deep learning and ensemble learning are powerful approaches to phishing detection because they all demonstrate higher accuracy, precision, better classification and possess the ability to learn from complex and more diverse datasets when compared to conventional approaches.

Finally, future research in phishing detection should integrate hybrid models such as deep learning and ensemble learning to address the above-mentioned limitations. As phishers employ more sophisticated techniques, there is a need for continuous model reforms to tackle the problem of phishing websites. Collaborative efforts between researchers and cybersecurity experts are also crucial in staying ahead of threats. When the right techniques are put in place, organizations can strengthen their defenses against phishing attacks, ensuring a safer digital environment for internet users.

References

1. Abdelhamid, N. (2014). Phishing detection based on associative classification data mining. *Expert Systems with Applications*, *59*, 59.
2. Abuzuraiq, A., Alkasassbeh, M., & Almseidin, M. (2020). Intelligent methods for accurately detecting phishing websites. *2020 11th International Conference on Information and Communication Systems (ICICS)*. https://doi.org/10.1109/ICICS49469.2020.239509
3. Adeojo, O. S., Egerson, D., Mewiya, G., & Edet, R. (2021). The ideology of baby-mama phenomenon: Assessing knowledge and perceptions among young people from educational institutions.
4. Ali, W. (2017). Phishing website detection based on supervised machine learning with wrapper features selection. *International Journal of Advanced Computer Science and Applications (IJACSA)*, *7*.
5. Basit, A., Zafar, M., Liu, X., Javed, A. R., Jalil, Z., & Kifayat, K. (2021). A comprehensive survey of AI-enabled phishing attacks detection techniques. *Telecommunication Systems*, *76*(1), 139–154. https://doi.org/10.1007/s11235-020-00733-2
6. Burbela, K. (2023). Model of detection of phishing URLs based on machine learning. *Faculty of Computing, Blekinge Institute of Technology, 371 79 Karlskrona, Sweden*, *40*.
7. Chicco, D., Jurman, G. The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification. BioData Mining 16, 4 (2023). https://doi.org/10.1186/s13040-023-00322-4
8. Daniel, M. A., Chong, S.-C., Chong, L.-Y., & Wee, K.-K. (2025). Optimizing phishing detection: A comparative analysis of machine learning methods with feature

selection. *Journal of Informatics and Web Engineering*, *4*(1), 200–212. https://doi.org/10.33093/jiwe.2025.4.1.15

9. Dutta, A. K. (2021). Detecting phishing websites using machine learning technique. *PLoS ONE*, *16*(17). https://api.semanticscholar.org/CorpusID:238635015
10. Dua, D., & Karra Taniskidou, E. (2017). *UCI Machine Learning Repository*. University of California, Irvine, School of Information and Computer Science. https://archive.ics.uci.edu/
11. Ejaz, A. (2023). Life-long phishing attack detection using continual learning. *Scientific Reports*, *13*.
12. Fadaei, S., Masoumi, B., & Abdi, F. (2020). Deep learning-based phishing detection using convolutional and recurrent neural networks. *Security and Privacy*, *3*(4), e109. https://doi.org/10.1002/spy2.109
13. Gillis, A. (2004). *What is cyber hygiene and why is it important?* TechTarget. https://www.techtarget.com/searchsecurity/definition/phishing
14. Gresele, L. (2023). Learning identifiable representations: Independent influences and multiple views. *[Journal Name Missing]*, *52*.
15. Guo, W., Wang, Q., Yue, H., Sun, H., & Hu, R. Q. (2025). *Efficient Phishing URL Detection Using Graph-based Machine Learning and Loopy Belief Propagation*. arXiv. https://arxiv.org/abs/2501.06912arXiv
16. Han, J., Shen, W., & Liu, X. (2022). Real-time threat intelligence sharing for phishing detection. *Cybersecurity*, *5*(1), 23.
17. Hannousse. A, Yahiouche. S., (2020). Towards Benchmark Datasets for Machine Learning Based Website Phishing Detection: An experimental study
18. Hannousse, A., & Yahiouche, S. (2021). Towards benchmark datasets for machine learning-based website phishing detection: An experimental study. *Engineering Applications of Artificial Intelligence*, *104*, 104347. https://doi.org/10.1016/j.engappai.2021.104347
19. Introducing Ultralytics "Confusion Matrix" https://www.ultralytics.com/glossary/confusion-matrix
20. Isik, O. (2024, October 29). Phishing attacks are evolving. Here's how to resist them. *Harvard Business Review*. Retrieved from https://hbr.org/2024/10/phishing-attacks-are-evolving-heres-how-to-resist-them?ab=HP-latest-text-8
21. Jerry, F. (1987). System security: A hacker's perspective. *INTEREX*
22. Khan, S. A., Khan, W., & Hussain, A. (2020). Phishing attacks and websites classification using machine learning and multiple datasets (A comparative analysis). *Intelligent Computing Methodologies: 16th International Conference, ICIC 2020, Bari, Italy, October 2–5, 2020, Proceedings, Part III* (pp. 301–313). Springer. https://doi.org/10.1007/978-3-030-60796-8_26
23. Khan.S, Khan.W, & Hussain. A, (2021). Phishing Attacks and Websites Classification Using Machine Learning and Multiple Datasets (A Comparative Analysis)
24. Kulkarni, A., Balachandran, V., Divakaran, D. M., & Das, T. (2024). Mitigating bias in machine learning models for phishing webpage detection. *arXiv preprint arXiv:2401.08363*. https://arxiv.org/abs/2401.08363
25. Mathews, L. (2017, May 5). Phishing scams cost American businesses half a billion dollars a year. *Forbes*. Retrieved from https://www.forbes.com
26. Micheal, C. (2024, November 21). Africa's top 4 countries with the highest scam losses in 2024. *Business Day*. Retrieved from https://businessday.ng/news/article/africas-top-4-countries-with-the-highest-scam-losses-in-2024/
27. Mohammed, R. M. (2014). Predicting phishing websites based on self-structuring neural network. *Neural Computing and Applications*, *443-458*.
28. Springbord. (2023, May 20). Data labeling in machine learning: Why is it important? *Springbord*. Retrieved from https://www.springbord.com/blog/data-labeling-in-machine-learning-why-is-it-important/
29. Oes, A., & Divakaran, D. (2022). Phishing detection leveraging machine learning and deep learning: A review. *Wireless Personal Communications*, *127*(3), 2663–2684. https://doi.org/10.1007/s11277-022-09994-3
30. Orugboh, O. G., Omabuwa, O. G., & Taiwo, O. S. (2024). Predicting Neighborhood Gentrification and Resident Displacement Using Machine Learning on Real Estate, Business, and Social Datasets. Journal of Social Sciences and Community Support, 1(2), 53-70.
31. Verma, R., & Das, A. (2018). What's in a URL? Understanding phishing URLs through lexical analysis. *Proceedings of the 2018 IEEE Symposium on Security and Privacy (SP)*, 1874–1890. https://doi.org/10.1109/SP.2018.00038
32. Weka Team. (2016). *Weka 3: Machine Learning Software in Java* (Version 3.8.6). The University of Waikato. https://www.cs.waikato.ac.nz/ml/weka/
33. 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.
34. Zhang, X., Li, W., & Zhang, Q. (2021). Phishing website detection using deep learning techniques: A survey. *Computers & Security*, *102*, 102152. https://doi.org/10.1016/j.cose.2020.102152
35. Zhang, Y., Hong, J., & Cranor, L. (2007). CANTINA+: A feature-rich machine learning framework for detecting phishing websites. *ACM Transactions on Information and System Security*, *14*(2), 1–28. https://doi.org/10.1145/2019599.2019603
36. Zhou, Y., Wang, H., & Lin, Z. (2023). Hybrid phishing detection models: A comparative analysis. *Journal of Cybersecurity and Digital Forensics*, *5*(1), 45–62.