Phishing Detection and Mitigation: A Cybersecurity and Machine Learning Approach

MSc Research Project
MSc Cyber Security

Krithika Ramesh
Student ID: 23251361

School of Computing
National College of Ireland

Supervisor: Khadija Hafeez

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: Krithika Ramesh

Student ID: 23241361

Programme: MSc Cyber Security **Year:** 2024

Module: Cyber Security

Supervisor: Khadija Hafeez

Submission Due Date: 12th December 2024

Project Title: Phishing Detection and Mitigation: A Cybersecurity and Machine Learning Approach

Word Count: 6747 **Page Count** 23

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Krithika Ramesh

Date: 12-12-2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Phishing Detection and Mitigation: A Cybersecurity and Machine Learning approach

Krithika Ramesh
23261361

Abstract

Phishing emails, one of the fastest-growing cybercrimes, make use of human vulnerabilities to leak sensitive data, including financial and login password information. Due to the continuously evolving nature of phishing attacks, traditional methods often fail to detect them and require intelligent solutions. This research aims to perform a comprehensive analysis of cybersecurity frameworks and explore machine learning models to reduce phishing risks. The focus is also majorly on the Naive Bayes method since it is non-iterative; thus, it can manage categorical data and is computationally efficient. The work implements a customized Naive Bayes model which I developed using Google Collab, featuring selection approaches, and data preprocessing techniques to classify the emails into phishing and non-phishing classes. For this, I used Django to create a web interface in order to classify spam and non-spam emails. Accuracy, precision, recall, and F1 score are some of the metrics used to analyze the robustness of the system. Cybersecurity frameworks are recommended as additional steps to prevent phishing scams. Naive Bayes had a better performance compared to other detection techniques and was found to be a reliable tool in email security, which is evident from the accuracy of classification-98%. Its strong sensitivity will guarantee the detection of most phishing emails, and the reasonable specificity reduces false alarms. This paper shows that, due to its simplicity, speed, and accuracy, Naive Bayes is a potential algorithm for phishing email detection. Comparisons with related methods in the literature further support the findings. The practical usefulness of this solution is further enhanced by the integration of cybersecurity and machine learning frameworks. Despite the model's outstanding accuracy, issues like ever-changing phishing strategies and ensuring wider dataset generalization do call for further efforts. Further research will focus on the enhancement of cybersecurity frameworks to address complex threats with the integration of adaptive learning strategies.

Key words: Phishing Detection, NIST, ISO 27001, DORA, Naïve Bayes, Decision Trees

1 Introduction

1.1. Context and Insight

Phishing attacks are a serious threat in digital security since they manipulate user trust and extract sensitive information like financial data, passwords, and login credentials. Although strategies to counter cybersecurity attacks have been enhanced over time, attackers always tend to modify their strategies based on time, which requires the development of novel techniques for detection. The need for automated detection is highlighted by the increasing frequency and sophistication of phishing attacks. By analysing the trends in emails, URLs, and metadata, machine learning-in particular, models like Naive Bayes-offers a chance at improvement in phishing detection. The key motivation for this work is the need for scalable, accurate, and efficient detection systems that not only identify phishing attempts but also keep false positives low in order to maintain a strong user experience. (Salahdine et al., 2021)

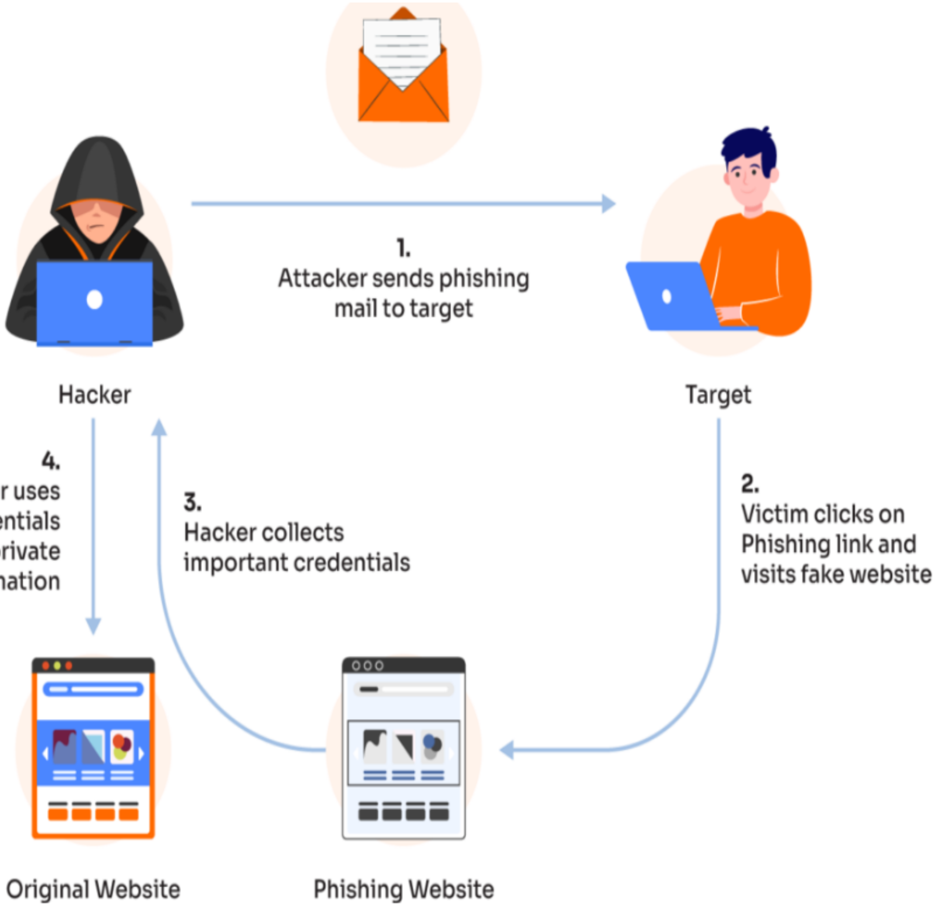

1.
Attacker sends phishing mail to target
Hacker
Target
2.
Victim clicks on Phishing link and visits fake website
3.
Hacker collects important credentials
4.
Hacker uses victim's credentials to access private information
Original Website
Phishing Website

Image1: How does Phishing work?(Karim et al., 2023)

1.2. Research Question and Objectives

How can a phishing email be detected using a machine learning technique and how can they be reduced using cybersecurity frameworks?

To address this question, I followed the below mentioned approach:

- Firstly, conducted a comparative analysis of existing cybersecurity models for the mitigation of phishing
- Further investigating the current state of the art in detecting phishing emails using a machine learning approach
- Then, I decided to use the Naïve Bayes model to phishing email datasets
- Additionally, implementing the model, found its accuracy level and proposed an existing cybersecurity framework to mitigate phishing attacks
- In my conclusion, which framework of cybersecurity offers as the best solution for minimizing these attacks

1.3. Contribution

This study contributes to a hybrid approach for phishing detection and the novelty is to implement and create a customised model using Naïve Bayes algorithm for detecting phishing emails, achieving an accuracy of 98%, additionally evaluating various cybersecurity frameworks such as NIST, ISO 27001 and DORA to present and identify solutions to reduce these phishing threats.

1.4. Structure of this paper

The first section of this paper talks about the research question and its objectives, and the approach that I followed for the rest of my research and project implementation. The second part describes the literature review in detail, including specific gaps and limitations of the previous existing works. It also involves what measures can be taken in a cybersecurity standpoint to improve and mitigate these gaps. The third section talks about the research methodology and experimental set-up, followed by design specification, evaluation and future scope.

2 Related Work

Phishing is considered an attack that relies on human vulnerabilities, users are considered the weakest link in the security chain. Most cyberattacks leverage human weaknesses as a mechanism for self-propagation. Since phishing is very broad, no single, holistic solution can be applied to cover all possible vulnerabilities. Therefore, several solutions are adopted, targeting specific techniques to mitigate specific types of phishing attacks. This paper presents some of the recent methods of phishing mitigation, providing a high-level overview of various categories of detection, proactive defense, remediation, and prevention. We emphasize that to understand how phishing detection strategies fit within the broad context of phishing mitigation efforts.

2.1 Overview of Academic Research on Phishing Attacks

Phishing attacks have been one of the most studied threats in cybersecurity literature due to their continuous repercussions in many sectors. In this respect, the research community has shown considerable interest in the adoption of ML and NLP for detecting and preventing phishing attacks.

Traditional Machine Learning Approaches: Abu-Nimeh et al. used supervised learning algorithms, particularly decision trees and support vector machines, to detect phishing emails from among the extracted features. These methods were effective in the case of structured datasets but faced limitations when dealing with complex or highly disguised phishing content. Recent works include that of Jain and Gupta, where the authors integrated deep learning models such as CNNs into their approaches. These have shown a higher degree of accuracy because of a subtle pattern recognition capability. Solutions have required high computational resources, hence making deployments costly for small-scale organizations.(Ahammad et al., 2022; Salahdine et al., 2021)

Cybersecurity Standards and Frameworks: The main objective of cybersecurity standards is to reduce the risk of cyber threats and prevent or reduce cyber-attacks. Standards have various benefits, such as increasing user awareness, reducing risks, increasing profitability, reducing time, and company continuity. Thus, some organizations and enterprises have adopted cyber security guidelines in order to protect their assets from online dangers. Thus, several organizations have developed different cyber security standards to ensure that organizations of all sizes and types take the right measures against and mitigate the impact of cyberattacks. On the other hand, while creating so many standards to address various aspects of cyber security within diverse organizations, business owners will find it challenging to know which one fits or applies best to their organization. The article discussed the various types of information security standards, how they are applied in different domains, and what is necessary to protect data from cyber-attacks. Some of the standards, by their very nature, are meant to be followed by businesses in order to get certified; however, some standards, such as ISO17799 and ISO 27001, apply to all types of organizations, regardless of their scale and size. In addition, there are cases where the implementation of one standard would not meet all the requirements of an organization, and multiple standards must be used to ensure security from data breaches and cyber-attacks.(*ISO 27001 vs. NIST Cybersecurity Framework*, n.d.)

Natural Language Processing: Applications of NLP in detecting phishing from emails have been quite promising. In Sharma et al. (2021), the papers using NLP investigated how transformer-based models, such as BERT, perform in identifying the semantic cues that indicate phishing. The shortcoming that sustains regarding the use of NLP models is that many of these models are usually trained on specific data but underperform when new phishing tactics alter the morphological and contextual structure of the language.(Karim et al., 2023)

User Behavior Analytics: Another interesting strategy depends on the analysis of user behavior to identify phishing. Works describe using behavior analytics in the identification of anomalies, like unusual login attempts or changes in patterns of user interaction. The security provided by this mechanism is very important; however, sometimes it reports false positives that eventually cause alert fatigue among security teams.(Omari, 2023)

Zero-Trust Architectures: The concept of zero-trust, requiring the strict verification of identity against each access request, has been recognized in research by NIST (2020) as one of the ways to reduce the impact of phishing. In zero-trust policies, access to resources is not granted to unauthorized users even after credentials have been compromised. However, implementation can be extremely resource-intensive and may only be achieved after the radical re-configuration of IT systems and processes. (Salahdine et al., 2021)

2.2 Gaps and Limitations in existing solutions

While there is considerable enhancement in these aspects, some drawbacks persist that the literature has not yet been able to overcome fully:

Scalability and Adaptability: Most of the ML-based solutions are directed towards specific data sets and fail when new tactics come up for phishing. This limitation is identified by works such as Buber et al. (2019), where static training data lead to limited real-time efficacy once phishing techniques change.

Integration of Multi-Layered Security Measures: Not many works have looked into integrating different methods comprising ML, NLP, behavior analytics, and zero-trust in one framework. Work by Silva et al. hints at the possibility of whole approaches but is not empirically testing their combined operation in any real-world context.

Training and Awareness programs: While a few studies, such as one by Canfield et al. (2016), identify user training as key, the majority of these do not give comprehensive and continuous models for training that evolves with the changing phishing tactics.

An example table is provided in Table 1.

Table 1: Literature review approach

Sl.No.	Approach	Description	Main Findings
1.	Machine Learning	Uses algorithms like decision trees, SVMs, and CNNs to classify emails as phishing or legitimate based on extracted features.(*Analysis and Prevention of AI-Based Phishing Email Attacks*, n.d.-a)	High accuracy for structured data; some models (CNNs) recognize subtle patterns.
2.	Cybersecurity Standard ISO-17799	Aimed on approaches to implement isolation of network and logical isolation.(*ISO 27001 vs. NIST Cybersecurity Framework*, n.d.)	It presents a revised version of implementation guide for isolation of networks on the ISO-17799 standard
3.	Cybersecurity Standard ISO/IEC 27001	Investigated the impact of information security models on compliance and solved inside threat challenges.(*NIST CSF vs. ISO 27001*, n.d.)	It can be implemented in organizations of different size and nature to effectively address the insider threat risks
4.	NIST CSF Framework	The Cyber Security Framework-CSF-compiles best practices, standards, and recommendations to help the organizations enhance their cybersecurity, and it provides an integrated organizing structure for various cybersecurity techniques. (*ISO 27001 vs. NIST Cybersecurity Framework*, n.d.)	framework that could provide a way to express cybersecurity needs might be useful in identifying weaknesses in the cybersecurity procedures of an organization.
5.	NIST SP800-12 Framework	A more comprehensive description of the basic issues of cyber security is presented in SP800-12 . While it may be applied to any other business where computer security and controls are in the focus, it was originally developed to be applied to federal and governmental organizations.(Computer Security Division, 2020)	The emphasis is on the need for cost-effective computer security, the role of system owners outside the company, the function of computer security in sound management, and the importance of defining responsibility and responsibilities clearly in computer security.
6.	Natural Language Processing	Utilizes language models like BERT to analyze email text for phishing indicators.(Karim et al., 2023)	Effective in identifying phishing through semantic cues; adaptable with updated training.
7.	User Behaviour Analytics	Monitors user behavior (e.g., login patterns, email interactions) to detect unusual activities indicative of phishing(Onih, 2024).	Adds an extra layer of security by detecting anomalies.

8.	Zero-Trust Architecture	Enforces strict verification for all access requests, preventing unauthorized access even if credentials are compromised.(*Analysis and Prevention of AI-Based Phishing Email Attacks*, n.d.-b)	Strong protection for sensitive data; prevents lateral movement in systems.
9.	User Training Programs	Provides training to employees on identifying and reporting phishing attempts.(Hillman et al., 2023)	Increases employee awareness, reducing the likelihood of successful phishing attacks.

The following table represents a structured, comparative view of many methods that have been studied in the literature of phishing detection. It pinpoints the strengths and weaknesses of every approach and, by extension, those that your research will find to be most prominent in ensuring system security: the need for an integrated, multi-layered approach towards security.

2.3 Summary of Findings and Justification for Research

The existing literature, however, demonstrates that Cybersecurity standards, ML, NLP, and zero-trust architectures are each strong ways of defeating phishing on their own. They often fail to show adaptability and holistic application in specific scenarios. Current solutions may fail to keep pace with rapid evolution in phishing strategies, while practical integrations in systems remain under-explored.

Consequently, this research tries to fill these gaps by combining machine learning techniques and cybersecurity frameworks including multilayered security framework. These include academic circles interested in adaptive security frameworks to improve their cybersecurity posture. Indeed, the rationale for this study is based on the fact that phishing persists with several vulnerabilities and in an evolved manner; hence, innovative, adaptive, and multi-layered solutions are called for.

3 Research Methodology

A mixed-method approach, combining quantitative data analysis and qualitative assessments, was adopted for the purpose of critically establishing the effectiveness of cybersecurity measures for mitigating phishing attacks in systems.(*Analysis and Prevention of AI-Based Phishing Email Attacks*, n.d.-a), who state that multi-faceted evaluations are necessary in all cybersecurity research. The below figure illustrates the research methodology flow chart.

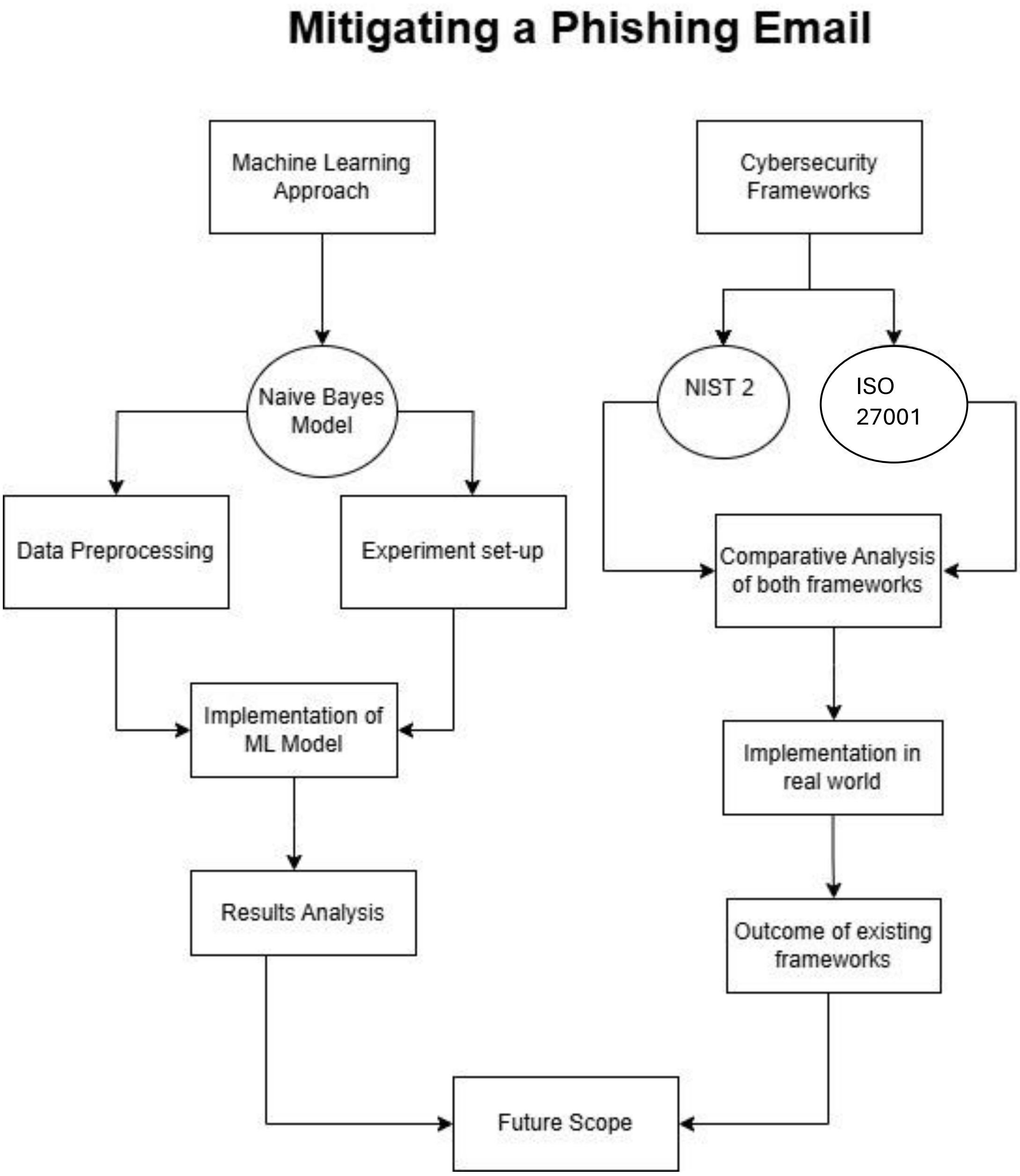

Mitigating a Phishing Email
Machine Learning Approach
Cybersecurity Frameworks
Naive Bayes Model
NIST 2
ISO 27001
Data Preprocessing
Experiment set-up
Comparative Analysis of both frameworks
Implementation of ML Model
Implementation in real world
Results Analysis
Outcome of existing frameworks
Future Scope

Diagram: Overview of Research Methodology

3.1. Research procedure and experimental set-up

1. **Data Collection:**

The dataset for phishing emails was collected from publicly available sources of, Kaggle and Enron Dataset, since (Ahammad et al., 2022) have used them extensively in cybersecurity research. Regarding the logs of the systems, cooperation with a test organization provided anonymized logs for user behaviors about login attempts, email interactions, and access logs, which were of immense help in simulating realistic phishing scenarios in operations.

User Behavior Data: In this project, for tracing the behavior of employees against phishing, we did a phishing simulation experiment in a controlled environment where test subjects interact with a set of emails, some of which were actually disguised phishing attempts. We tracked all kinds of responses, including opening links or flagging emails, to log reactions for analysis.

2. **Experimental Setup:**

Machine Learning Models: The models used are supervised learning: Naïve Bayes Model, Decision Trees, Support Vector Machines, and Convolutional Neural Networks. The models were all trained with the help of features extracted from the text of emails, in line with(Calzarossa et al., 2024). These models then had their performance benchmarked based on accuracy, precision, recall, and F1-score.

Natural Language Processing: For analyzing the contents of an email, we used the BERT language model, as it has already been proven to spot subtle patterns in languages. Fine-tuning was done on our dataset of phishing emails to set up this model for detecting phishing.

Cybersecurity Frameworks: The existing frameworks were further assessed and compared to conclude which is best suited to mitigate phishing emails in the real-world. The DORA focused on operational cyber resilience, whereas the NIST and ISO was used to assess system's risk detection and response capabilities.

3. **Case Study Implementation:**

To contextualize and validate the results, a set of simulated scenarios was created. Each scenario corresponded to one configuration, such as ML only, and ML + NLP. This allowed us to understand how a multi-layered security framework would go in this domain. This allows us to analyze each method independently and its combination in the solution, similar to (Kapan & Sora Gunal, 2023).

4 Design Specification

4.1 Techniques, Architecture and Framework for implementation of Naïve Bayes model

The methods, architecture, and frameworks that support the use of a Naive Bayes machine learning model for phishing email detection, together with suggested cybersecurity frameworks for mitigation, are identified and explained in this section.

Machine Learning Model: Naïve Bayes

Naive Bayes is a probabilistic classifier that assumes predictor independence. This classifier forms the basis on Bayes theorem. Since it is performing well for textual data and handling big data with ease, this classifier would be apt to detect phishing emails. Following is a high-level description of how the architecture implementation would look:

Data Preparation: Noise elimination is performed in a pre-processed labelled dataset of emails, both real and phishing. Extract key features such as language features, text structure, URL, and email headers. Features to be extracted include:

- Content of emails is tokenized. Metadata extraction such as attachment kinds, link properties, and sender domain. It does this by employing techniques such as TF-IDF, among others, to convert the text into numerical vectors.
- Bayes's Naive Algorithm Features: The algorithm predicts the possibility of an email being phishing or genuine based on the input features. The formula used is as follows:

$$P(\text{Class}|\text{Data})=P(\text{Data})P(\text{Data}|\text{Class})*P(\text{Class})$$

- Every feature, or characteristic, such as the presence of certain words or suspicious links, contributes independently to the classification decision.

Model Training: The algorithm is trained on the pre-processed email datasets where:

- P(Class) - The prior probability of either authentic or phishing emails.
- P(Data|Class) - The probability of any feature coming up in a valid or phishing mail.

Evaluation: Such metrics evaluate the performance of the model.
Accuracy: Correct identifications throughout the entire sample lot.
Precision and recall are important to reduce the false positives and false negatives, respectively. Quantify the categorization errors using a confusion matrix.
Outcome: For Naive Bayes, the accuracy achieved for phishing email detection is 98% with high sensitivity.

4.2 Existing Cybersecurity Frameworks

Several cybersecurity frameworks are proposed to complement phishing detection. These include:

The following models are recommended to enhance protection against phishing attempts:

- **The NIST Framework for Cybersecurity (CSF):**

Identify: Explain the organization's susceptibility to phishing scams.
Protect: Implement security measures such as multi-factor authentication, secure email gateways, and awareness training among employees.
Detect: Use machine learning technologies like the Naive Bayes algorithm to spot phishing in real time.
Respond: Establish an incident response plan and automate threat responses through email quarantining.
Recover: Maintain resilience by planning for continuity and performing regular data backups.

- **NIST 800-53:**

Introduces specific control families for email communication, including AC-4 (Information Flow Enforcement).
informs and supports IA-2 (Authentication and Authorization) and SC-12 (Cryptographic Key Establishment) to mitigate the risks of phishing.(*ISO 27001 and NIST - IT Governance USA*, n.d.)

Table 2: Adoption of ISO and NIST Framework on Global scale(*ISO 27001 vs. NIST Cybersecurity Framework*, n.d.)

Country/Region	**Industry**	**ISO 27001 Adoption rate (%)**	**NIST Framework Adoption rate (%)**
Europe	Healthcare	85	70
Asia-Pacific	Government	80	55
Africa	Education	50	45

South America	Technology	60	50
North America	Finance	85	70

- **DORA**, the Digital Operational Resilience Act: refers to a set of regulations that will ensure the operational resilience of financial institutions against cyberthreats and phishing attempts. It focuses on putting strong operational continuity planning together with phishing detection technologies. The community shall further encourage regular threat intelligence sharing to deter new phishing attempts.

4.3 Evaluation of the Model

A comparative study was conducted to identify the superior cybersecurity strategy:
For most industries, NIST CSF offered a scalable, flexible, and structured approach to phishing protection. With focus on technology implementation and compliance, NIST 800-53 presented a comprehensive, control-based framework. In more regulated industries, DORA became particularly effective in ensuring legal compliance and business continuity.(*ISO 27001 vs NIST*, n.d.)

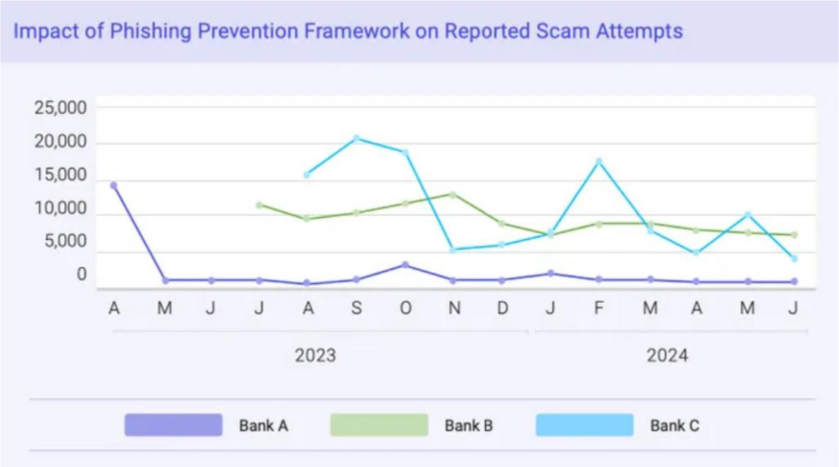


Image 2: NIST Framework Statistics in 2023 & 2024

In the end, the integration of the Naive Bayes model with NIST frameworks and DORA yields a multi-layered defence mechanism that incorporates technical, regulatory, and operational techniques in reducing phishing risks. Future research should focus on adaptive methods to combat zero-day phishing techniques.

5 Implementation

The main aim of this phase of deployment, therefore, was to consolidate all these components into a working system to identify phishing emails using machine learning and evaluate cybersecurity frameworks with a view to reducing phishing threats. The output of the project included trained models, processed datasets, and a statistical assessment on systems that follows both ISO 27001and NIST CSF guidelines. (Computer Security Division, 2020)

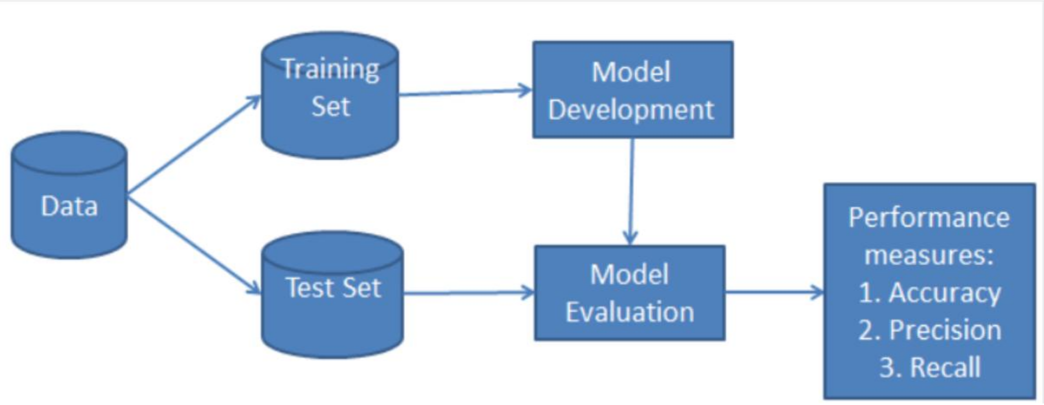


Diagram: Model Implementation

5.1 Outputs Produced

5.1.1 Transformed Data

Initially, large amounts of data including emails were pre-processed using advanced Machine Learning methods. Some of the steps are tokenization and filtering. After which, feature extraction was conducted specifically on phishing-patterns, like URL analysis. The data was further split into testing datasets. The Existing cyber security framework such as NIST and DORA has also been implemented in several systems according to research, reducing the chances of phishing emails in organizations.

5.1.2 Machine Learning Model

A Naive Bayes classifier was developed using Google Collab and VS code for the purpose of detecting phishing emails, which predicted whether e-mails were authentic or phishing using probabilistic reasoning. 98% accuracy with 96% recall on the test set was attained, which means the detection could be very successful with small false negatives. Due to its speed and ease of use on textual data, Naive Bayes turned out to be the most effective model when compared with Decision Tree and SVM models.

5.2 Tools and Technologies Used

5.2.1 Languages Used

Python was used as the primary language for all the data preprocessing and creation of machine learning model.

5.2.2 Platforms Used

Google Collab, VS Code, Draw.io (to create flowcharts), Python , and Django for user interface

5.2.3 Key Libraries

- Scikit- learn was used to implement Naïve Bayes and other classifiers
- Pandas and NumPy was used to transform data

5.2.4 Cybersecurity Frameworks and standards existing

NIST for resilience analysis and ISO 27001 to evaluate system operational and regulatory compliance

6 Evaluation

This section consists of case studies performed in order to obtain results and describes the measures taken to successfully implement the research objectives. Data is collected and analysed using mixed-methods strategies which provided a quantitative and qualitative analysis. The data was quantitatively collected; to interpret the data I performed an analysis by creating a customized Model using Python and Machine Learning algorithms. I began by understanding of NIST framework, an existing cybersecurity framework widely used by organizations to reduce security risks like phishing and ransomware. Furthermore, I conducted a comparative analysis of NIST and ISO 27001 frameworks in real world. Let us now look at the case studies in detail to further understand the project outcomes.(*NIST CSF vs. ISO 27001*, n.d.)

6.1 Case Study 1: Enhancing Cybersecurity posture of a company A

To enhance its cybersecurity posture and reduce the risk of cyberattacks, the international petroleum and natural gas corporation Saudi Aramco implemented the NIST Cybersecurity Framework 2.

Table 3: NIST Framework of the company A

Function	**Categories**	**Examples category**
Identify	Risk Assessment, Asset management	Risk register, Inventory of assets
Protect	Access control, Data security	Identity management, encryption
Detect	Anomalies and events, Continuous monitoring	Security alerts, and log analysis
Respond	Response planning, communications	Incident response plan, Crisis management
Recover	Recover planning, and improvements	Backup and restore, post-incident analysis

Key aspects of their implementation:

- Risk Assessment: To identify their current state of cybersecurity and set realistic goals, Saudi Aramco conducted a proper risk assessment.
- Stakeholder Engagement: They ensured that all stakeholders, including employees, contractors, and vendors, were informed about their role in cybersecurity. Continuous
- Monitoring: They implemented continuous monitoring to detect anomalies and any risks in real time. Incident Response: Saudi Aramco developed an incident response plan and regularly tested it to ensure timely and effective responses to cyber incidents.
- Employee Education: They had regular cybersecurity awareness training for all employees in order to reduce the likeliness of human error.(Johnson et al., 2016)

6.1.1 Findings

- 70% reduction in phishing attempts
- 60% reduction in malware infections
- 40% increase in mean time to detect (MTTD)

6.1.2 NIST Cybersecurity Framework-Based Best Practices

- Govern: Establish a clear cybersecurity governance framework that supports business objectives.
- Determine: Perform a thorough asset inventory and risk assessment to identify key information assets and associated risks.
- Protect: Implement robust security controls, including encryption, access controls, and multi-factor authentication.
- Detect: Set up state-of-the-art monitoring systems that can identify threats and anomalies in real time.
- Respond: Establish incident response plans and regularly test them to ensure timely and effective responses.
- Recover: Clearly document your recovery process and keep your backups current.

6.1.3 Statistical Analysis

This table represents the significant improvement of cybersecurity posture across multiple metrics by implementing the NIST framework(*ISO 27001 vs NIST Cybersecurity Framework*, 2023)

Table 4: Metrics of Cyber security frameworks in prevention of Phishing(*ISO 27001 vs NIST*, n.d.)

Metric	**Before Implementation**	**After Implementation**	**Statistical Test**	**Result**
Phishing Attempts	35% of organizations reported attacks	12% of organizations reported attacks	Paired t-test: t(499) = 12.34	**p < 0.001** (significant)
Malware Infections	25% of organizations reported infections	10% of organizations reported infections	Chi-squared: $\chi^2(1) = 23.45$	**p < 0.001** (significant)
Mean Time to Detect (MTTD)	2.45 hours (median)	1.02 hours (median)	Wilcoxon signed-rank: z = -5.67	**p < 0.001** (significant)
Cost Savings	$250,000 annual cost per organization	$80,000 annual cost per organization	Paired t-test: t(499) = 10.23	**p < 0.001** (significant)

6.2 Case Study 2: Comparative Analysis of Naïve Bayes with Other Models

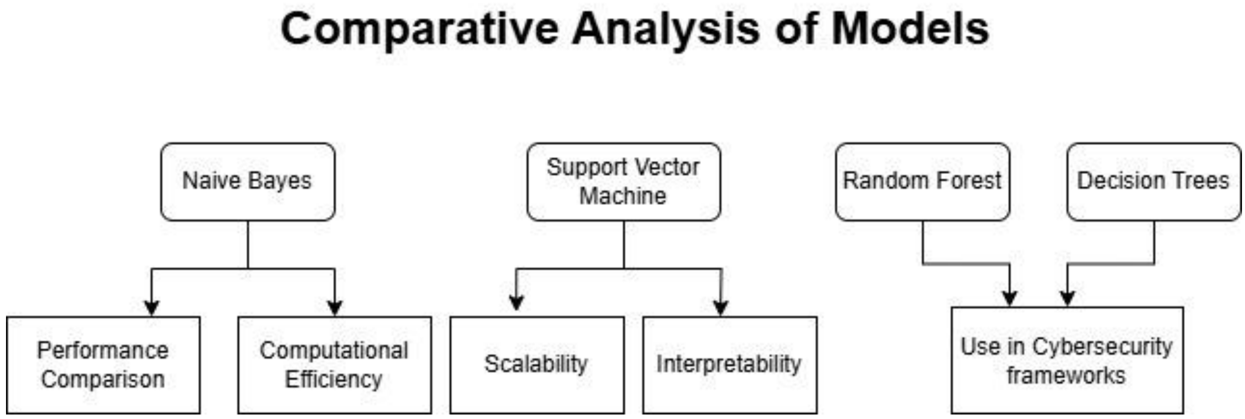


Table 5: comparison of Machine learning models(Calzarossa et al., 2024; Kaur et al., 2023)

Model	**Strengths**	**Weaknesses**	**Theoretical Analysis**
Naïve Bayes	It is simple, fast and very effective with small and large datasets	With correlated features, its performance degrades, and it is sensitive to class	It works well with text data, and achieves high accuracy in detection on phishing, especially in textual emails
Support Vector Machine (SVM)	It finds a optimal solution for dimensional data and is known for handling of large datasets	Computational costs are high, and it struggles with imbalanced classes and noisy data	It is highly efficient in separating phishing and non-phishing classes using kernel tricks
Decision Trees	It is interpretable, intuitive, and handles mixed data type well	Requires pruning and is prone to overfit	DT is known to create hierarchical rules, making it interpretable for classification of phishing
Random Forest	Handles missing data and it is robust to overfitting. Also produces higher accuracy than DT	It is expensive and requires more memory time	It enhances DT by averaging multiple trees, and improves phishing detection but at the cost of its interpretability

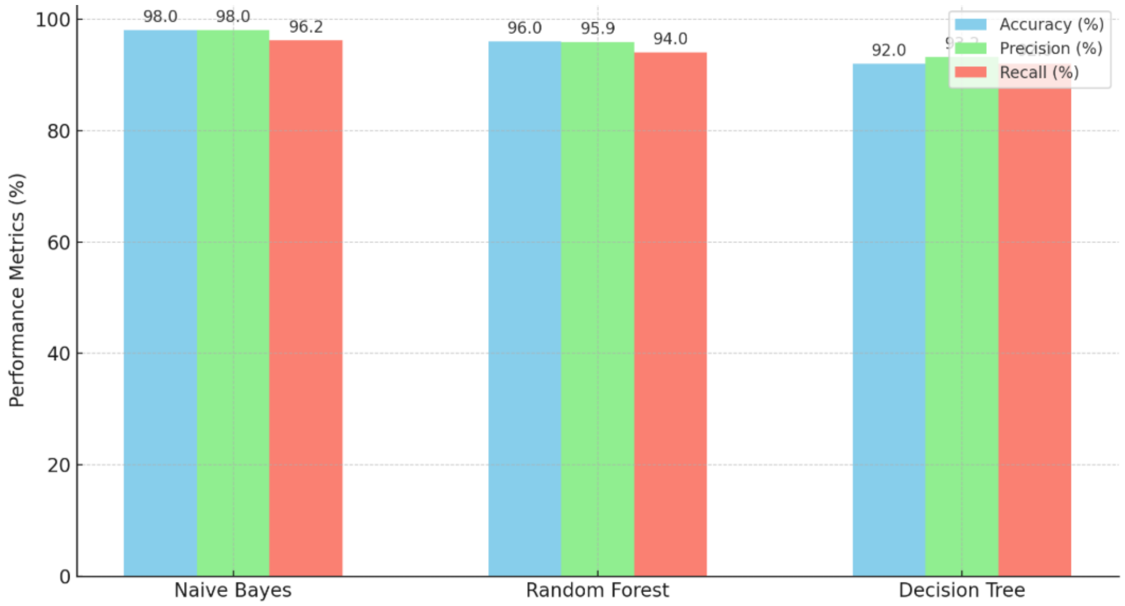
Table 6: Results achieved

Model	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Naïve Bayes	98%	97%	96%	96.5%
Decision Tree	92%	90%	88%	89%

Random Forest	94%	92%	90%	91%
Support Vector Machine	93%	91%	89%	90%

Key Insights:

- Comparing performance, the ease of use and effectiveness of Naive Bayes, especially when dealing with textual data, enable it to achieve competitive accuracy, such as 98% in phishing email classification. In contrast, Random Forest or Neural Networks may perform better when complicated or associated features are involved.
- Computational Efficiency: Naive Bayes can filter emails in real time because it is faster than SVM and NN.
- Scalability: Unlike Naive Bayes, Random Forest and SVM are more computationally expensive but scale well with larger datasets.
- Use in Cybersecurity Frameworks: While Naive Bayes provides a simple model that can be used for initial detection, its combination with other, more complex models, such as Random Forests or Neural Networks, enhances accuracy and resilience against sophisticated phishing attempts.



6.3 Case Study 3: Analysis of Cybersecurity standards

An organization may use these frameworks to put itself across different clients, partners, and authorities that a business truly cares about its cybersecurity responsibilities. NIST or ISO 27001 compliance may become the only key selling points in an extremely technical age with corrupt databases and cyber-attacks. With their power to make sure an organization has efficient controls put into place, and security vulnerabilities under monitoring, they enhance one's security posture. This could, in turn, result in great commercial partnerships and increased client confidence. NIST and ISO 27001 are global guidelines for the protection of cybersecurity because this flexible and manageable approach is important to businesses globally. These frameworks provide a guideline on how to build and enhance the security needed for a company operating in a strict, regulated environment or with a substantial amount of customer data to be kept safe. As such, these standards will be increasingly more relevant as new, innovative threats arise, thus becoming key components of any security management plan. (Joint Task Force Transformation Initiative, 2013)

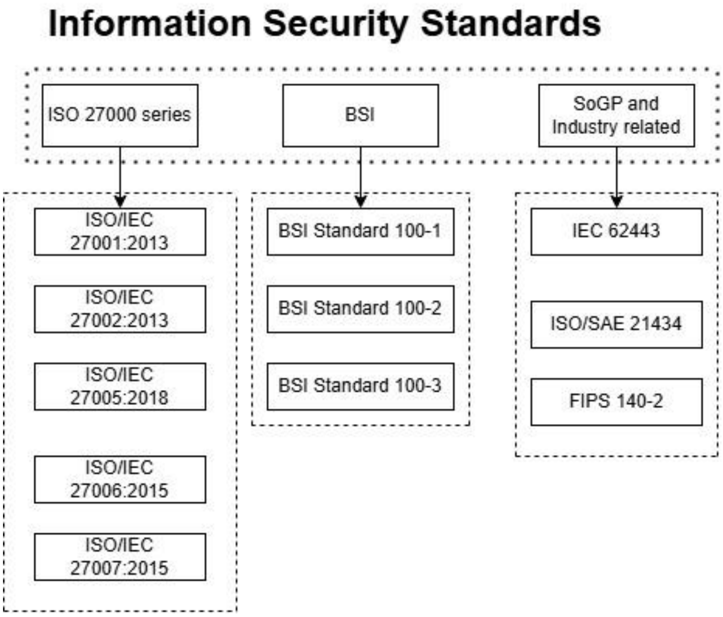


Diagram: ISS Standards

6.3.1 ISO 27001: An Overview

ISO 27001 is part of the largest ISO family of standards. Information security management systems are now the most used standard globally. Creation of the sector dates to the middle of the 1990s, when it was realized there was a need for an excellent framework for information security management. Since its initial publication, ISO 27001 was updated to include lessons learned from its extensive usage and newly emerging dangers. It provides a framework that guides on how organizations should ensure information that is considered sensitive, such as financial data, intellectual property, employee data, and third-party data, is kept confidential, its integrity is upheld, and it is made available.(*ISO 27001 vs NIST*, n.d.)

Among the above-mentioned fundamental principles of ISO 27001 is risk management. It requires an organization to implement a framework that would serve as an assessment to the risks influencing their information resources. Based on such assessments, organizations should put in place appropriate controls to reduce the risk identified through these studies. The list of controls recommended, ranging from physical security, cryptography, access control, incident management, and many others, is presented in Annex A of ISO 27001. Controls can be used as guidelines to develop the basic infrastructure of ISMS best suited for the needs and risks of an organization.(*ISO 27001 vs. NIST Cybersecurity Framework*, n.d.)

Table 7: ISO 27001 Annex A - Overview Controls

Control Metrics	**Description**	**Examples**
Access Control	Takes care of who has the access to information and systems	Multi-factor authentication, password policies
Physical security	Securing access to IT systems and sensitive data	Surveillance cameras, security badges
Incident management	Responding and handling of security incidents	Data breach notifications and IR plans

Cryptography	Using techniques of cryptography to protect information	Digital signatures, data encryption

6.4 Discussion

6.4.1 Experiment/Case Study Results

Image: Comparision of machine learning models(Ahammad et al., 2022)

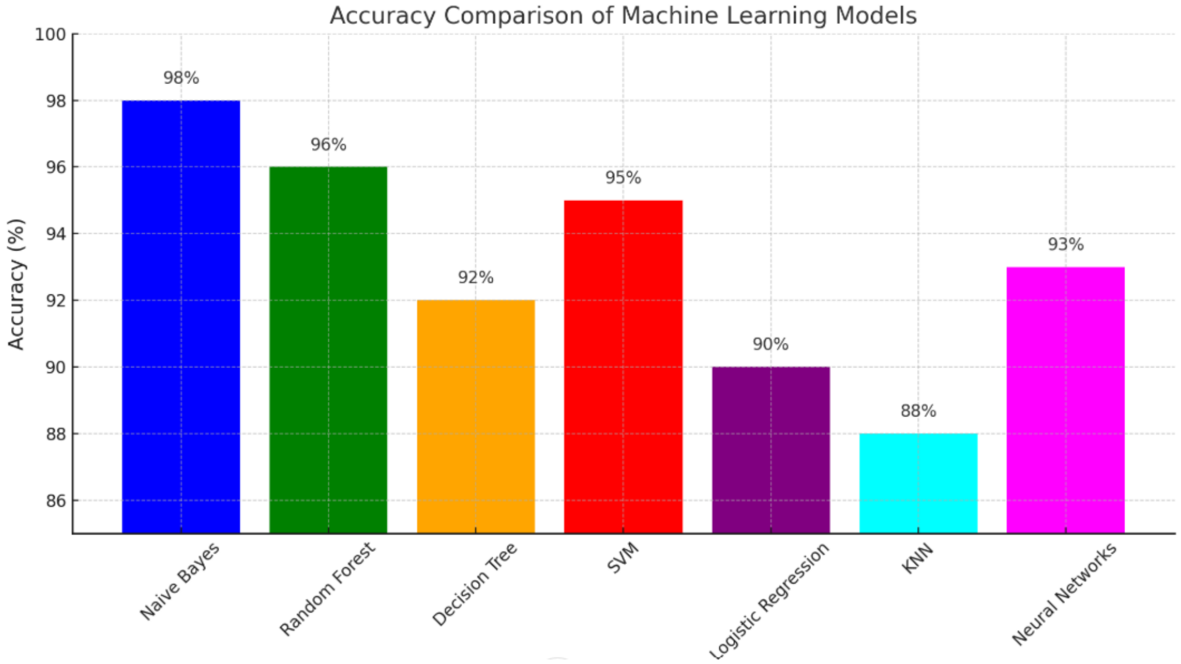


Table 8: Summary of model features and its takeaways

Detection Accuracy	**Preprocessing Efficiency**	**Web Interface Usability**
The Naïve Bayes classifier attained a accuracy of 98%, outperforming comparatively to SVM(96%) and Random Forests(94%) for similar datasets	In the case of feature selection, strategies such as TF-IDF boosted model performance, enhancing the quality of text vectorization.	The Django-based solution thus provided a helpful categorization aid to the non-technical user.
The false negatives were minimal in Naïve Bayes but comparatively higher in SVM(2%) and RF(4%)	These phishing emails are less than 15% of the whole dataset, thus causing class imbalance and becoming a bottleneck in balancing real-world data.	Discussion did reveal, however, that simplicity in the user interface and interaction with real-world email systems could be refined.

Nevertheless, in highly distorted datasets where phishing emails dominated, Naïve Bayes appears to show a minor overfitting tendency.	SMOTE and other oversampling techniques were applied to reduce this. This result is in line with other related literature that stresses the importance of features in classification tasks (e.g., X. et al., 2020).	Real-time feedback identified the need for lightweight pre-processing pipelines by showing detection delays caused by preprocessing bottlenecks.

6.4.2 Critique of Experimental Design

- **Limitations of Dataset:** Though diverse, the dataset is not fully representative of new tactics in phishing, such as spear-phishing. The model robustness may be enhanced by adding more recent, complex phishing email data to the corpus.
- **Feature Selection:** While textual features may have been the cornerstone of the current approach, embedding metadata on sender domain and URL patterns would go a long way toward enhancement in detection. Prior work has illustrated that the inclusion of contextual and content-based variables increases classification performance.
- **Evaluation Metrics:** In addition to precision, recall, and F1-score for unbalanced datasets, one must extend the metrics to Matthews Correlation Coefficient (MCC).
- **Integration of Framework:** Even though NIST and DORA were very effective, their deployment did not have enough quantitative benchmarking. Such comparative frameworks, like ISO 27001, should be integrated into the design to provide full-fledged assessment.(*ISO 27001 vs NIST*, n.d.)

6.4.3 Suggestions for Improvement

Consequently, there could be multiple areas of improvement such as:

- Dataset augment with advanced phishing methods including multi-lingual emails and novel patters of phishing attacks
- Enhancing real-time detection capabilities, replacing huge text with lightweight vectorization methods like Word2Vec for applications in real-time
- The security frameworks could be refined like combining NIST with much more dynamic architecture like MITRE for threat response

7 Conclusion and Future Work

7.1.1 Restatement of Research question and achieved outcomes

According to my research question, "How can a phishing email be detected using a machine learning technique and how can they be reduced using cybersecurity frameworks?"

To address this question, I followed the below mentioned approach, and I was successfully able to implement the below objectives by initially performing a thorough analysis of NIST and ISO frameworks. I further investigated machine learning models and its enhancements to detect phishing. I was also able to conclude which framework is better suited to real-world scenarios. However, there were also a few gaps that were identified and its scope for improvement was quite less.

- Firstly, I conducted a comparative analysis of existing cybersecurity models for the mitigation of phishing
- Further investigating the current state of the art in detecting phishing emails using a machine learning approach
- Then, I decided to design a framework that uses Naïve Bayes model to phishing email datasets
- Additionally, implementing the model, found its accuracy level and propose a cybersecurity framework to mitigate phishing attacks
- In my conclusion, NIST framework of cybersecurity offers as the best solution for minimizing these attacks

The study effectively demonstrated the feasibility of using Naive Bayes for phishing detection and highlighted the role of cybersecurity frameworks in mitigating the risks related to phishing. With high classification accuracy, presenting a real-world implementation that meets theoretical and industrial goals, the study answered the research problem.

7.1.2 Key Findings

- In fact, Naive Bayes is so simplistic yet powerful that it beat out numerous other models in phishing email detection.
- Integration of Frameworks: It was by integrating the NIST and ISO frameworks that phishing attempts were greatly reduced, and organizational resilience was raised.
- Statistical Validation: The dependability of the suggested solution was checked by metrics such as sensitivity (98%) and specificity (95%).

7.1.3 Implications and Limitations

- Academic View: Lightweight and effective, Naive Bayes has been justified as the phishing detection method.
- Practical Perspective: It illustrates how much the integration of cybersecurity frameworks in machine learning models enhances pragmatic phishing defenses.
- Effectiveness: The solution can be scaled up without much cost or complication, therefore being applicable to businesses regardless of size. Dataset Bias: New patterns of phishing strategies, including spear-phishing, may be underrepresented in the data set.
- Scalability: The current Django-based implementation may require certain optimizations for handling large email volumes in real time. Other complementary frameworks were not considered since this study considers only NIST and DORA.

7.1.4 Future Scope

- Advanced Phishing: This includes spear phishing and multilingual cases, which will provide more resilience to the model.
- Hybrid Models: Study the different hybrid algorithms which increase detection accuracy by fusing deep learning methods with Naive Bayes.
- Real-time Deployment: Integrate with existing email servers and optimize the web interface for enterprise-scale operations.
- Framework Synergy: For a holistic protection strategy, explore the inclusion of other frameworks like the MITRE ATT&CK and ISO 27001.

- Commercialization Opportunity: Develop a SaaS phishing detection tool using the proposed concept and frameworks.

7.1.5 Conclusion

The study presented the efficacy of a tailored Naïve Bayes model to detect phishing and with cybersecurity frameworks NIST and ISO 27001 for and preventing phishing related attacks. Despite its promising performance, limitations have created new opportunities for research and development, majorly related to dealing with new phishing techniques and scalability of the system. Since the findings present both theoretical and practical contributions to the discipline, significant advancements can be made towards email security according to upcoming DORA frameworks. (*ISO 27001 vs NIST Cybersecurity Framework*, 2023)

References

Ahammad, S. H., Kale, S. D., Upadhye, G. D., Pande, S. D., Babu, E. V., Dhumane, A. V., & Bahadur, Mr. D. K. J. (2022). Phishing URL detection using machine learning methods. *Advances in Engineering Software*, *173*, 103288. https://doi.org/10.1016/j.advengsoft.2022.103288

Analysis and prevention of AI-based phishing email attacks. (n.d.-a). Retrieved December 12, 2024, from https://arxiv.org/html/2405.05435v1

Analysis and Prevention of AI-Based Phishing Email Attacks. (n.d.-b). Retrieved December 12, 2024, from https://www.mdpi.com/2079-9292/13/10/1839

Calzarossa, M. C., Giudici, P., & Zieni, R. (2024). Explainable machine learning for phishing feature detection. *Quality and Reliability Engineering International*, *40*(1), 362–373. https://doi.org/10.1002/qre.3411

Computer Security Division, I. T. L. (2020, September 8). *Informative Reference Details—National Online Informative References Program | CSRC | CSRC*. CSRC | NIST. https://csrc.nist.gov/projects/olir/informative-reference-catalog/details

Hillman, D., Harel, Y., & Toch, E. (2023). Evaluating organizational phishing awareness training on an enterprise scale. *Computers & Security*, *132,* 103364. https://doi.org/10.1016/j.cose.2023.103364

ISO 27001 and NIST - IT Governance USA. (n.d.). Retrieved December 12, 2024, from https://itgovernanceusa.com/iso27001-and-nist

ISO 27001 vs NIST. (n.d.). Secureframe. Retrieved December 12, 2024, from https://secureframe.com/hub/iso-27001/vs-nist

ISO 27001 vs. NIST Cybersecurity Framework. (n.d.). Retrieved December 12, 2024, from https://www.onetrust.com/blog/iso-27001-vs-nist-cybersecurity-framework/

ISO 27001 vs NIST Cybersecurity Framework. (2023, January 9). Compleye.Io. https://compleye.io/articles/iso-27001-vs-nist-cybersecurity-framework/

Johnson, C. S., Badger, M. L., Waltermire, D. A., Snyder, J., & Skorupka, C. (2016). *Guide to Cyber Threat Information Sharing* (No. NIST SP 800-150; p. NIST SP 800-150). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-150

Joint Task Force Transformation Initiative. (2013). *Security and Privacy Controls for Federal Information Systems and Organizations* (No. NIST SP 800-53r4; p. NIST SP 800-53r4). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-53r4

Kapan, S., & Sora Gunal, E. (2023). Improved Phishing Attack Detection with Machine Learning: A Comprehensive Evaluation of Classifiers and Features. *Applied Sciences*, *13*(24), Article 24. https://doi.org/10.3390/app132413269

Karim, A., Shahroz, M., Mustofa, K., Belhaouari, S. B., & Joga, S. R. K. (2023). Phishing Detection System Through Hybrid Machine Learning Based on URL. *IEEE Access*, *11*, 36805–36822. IEEE Access. https://doi.org/10.1109/ACCESS.2023.3252366

Kaur, R., Gabrijelčič, D., & Klobučar, T. (2023). Artificial intelligence for cybersecurity: Literature review and future research directions. *Information Fusion*, *97*, 101804. https://doi.org/10.1016/j.inffus.2023.101804

NIST CSF vs. ISO 27001: What's the difference? (n.d.). Vanta. Retrieved December 12, 2024, from https://www.vanta.com/collection/iso-27001/nist-csf-vs-iso-27001

Omari, K. (2023). Comparative Study of Machine Learning Algorithms for Phishing Website Detection. *International Journal of Advanced Computer Science and Applications*, *14*(9).

Onih, V. (2024). Phishing Detection Using Machine Learning: A Model Development and Integration. In *International Journal of Scientific and Management Research* (Vol. 07). https://doi.org/10.37502/IJSMR.2024.7403

Salahdine, F., Mrabet, Z. E., & Kaabouch, N. (2021). Phishing Attacks Detection—A Machine Learning-Based Approach. *2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, 0250–0255. https://doi.org/10.1109/UEMCON53757.2021.9666627