# HYBRID MACHINE LEARNING MODEL TO DETECT AND MITIGATE EMAIL PHISHING ATTACKS

## A MINOR PROJECT REPORT

**Submitted by**

AKBARSAGARI MOHMMAD FAZIL[RA2111030010175]

PALLAPOLU SAI VARDHAN REDDY[RA2111030010186]

JAINI ESWAR[RA2111030010190]

**Under the guidance of**

DR.S. PRABAKERAN

(Associate Professor, Department of NWC)

in partial fulfilment for the award of the degree of

## BACHELOR OF TECHNOLOGY
## COMPUTER SCIENCE & ENGINEERING

**DEPARTMENT OF COMPUTATIONAL INTELLIGENCE**

**FACULTY OF ENGINEERING AND TECHNOLOGY**

**SRM INSTITUTE OF SCIENCE AND TECHNOLOGY KATTANKULATHUR-603203**

**NOVEMBER 2023**

# Department of Computational Intelligence

## SRM Institute of Science & Technology

## Own Work* Declaration Form

**This sheet must be filled in (each box ticked to show that the condition has been met). It must be signed and dated along with your student registration number and included with all assignments you submit – work will not be marked unless this is done.**

**To be completed by the student for all assessments**

**Degree/ Course : Bachelor of technology in computer science and Engineering with a Specialization in Cyber Security**

**Student Name : AKBARSAGARI MOHMMAD FAZIL,PALLAPOLU SAI VARDHAN REDDY, JAINI ESWAR**

**Registration Number : RA2111030010175, RA2111030010186, RA2111030010190**

**Title of Work : HYBRID MACHINE LEARNING MODEL TO DETECT AND MITIGATE EMAIL PHISHING ATTACKS**

**We hereby certify that this assessment compiles with the University's Rules and Regulations relating to Academic misconduct and plagiarism**, as listed in the University Website, Regulations, and the Education Committee guidelines.**

**We confirm that all the work contained in this assessment is my / our own except where indicated, and that I / We have met the following conditions:**

- **Clearly referenced / listed all sources as appropriate**
- **Referenced and put in inverted commas all quoted text (from books, web, etc)**
- **Given the sources of all pictures, data etc. that are not my own**
- **Not made any use of the report(s) or essay(s) of any other student(s) either past or present**
- **Acknowledged in appropriate places any help that I have received from others (e.g. fellow students, technicians, statisticians, external sources)**
- **Compiled with any other plagiarism criteria specified in the Course handbook / University website**

**We understand that any false claim for this work will be penalized in accordance with the University policies and regulations.**

| |
|---|
| **DECLARATION:** |
| I am aware of and understand the University's policy on Academic misconduct and plagiarism and I certify that this assessment is my / our own work, except where indicated by referring, and that I have followed the good academic practices noted above. |
| RA2111030010175<br>RA2111030010186<br>RA2111030010190 |

# ACKNOWLEDGEMENTS

We express our humble gratitude to **Dr. C. Muthamizhchelvan**, Vice-Chancellor, SRM InstituteofScience and Technology, for the facilities extended for the project work and his continued support.

We extend our sincere thanks to **Dr. T. V. Gopal ,** Dean-CET, SRM Institute of Science and Technology, for his invaluable support.

We wish to thank **Dr. Revathi Venkataraman**, Professor and Chairperson, School of Com puting, SRM Institute of Science and Technology, for her support throughout the project work.

We encompass our sincere thanks to , **Dr. M. Pushpalatha**, Professor and Associate Chairperson, School of Computing and **Dr. C.Lakshmi,** Professor and Associate Chairperson, School of Computing, SRM Institute of Science and Technology, for their invaluable support.

We are incredibly grateful to our Head of the Department, **Dr. M. Lakshmi**, Professor, Department of Computational Intelligence, SRM Institute of Science and Technology, for her suggestions and encouragement at all the stages of the project work.

We want to convey our thanks to our Project Coordinators Panel Head, Dr .S. Prabakaran, Associate Professor and Panel Members Dr .Sundarrajan, Assistant Professor, Dr. M Mahalakshimi Assistant Professor,Department of Networking and Communications SRM Institute of Science and Technology, for their inputs during the project reviews and support.

We register our immeasurable thanks to our Faculty Advisor, Dr A Arun Department Of Networking and Communication, SRM Institute of Science and Technology, for leading and helping us to complete our course.

Our inexpressible respect and thanks to our guide, Dr S. Prabakaran Department Of Networking and Communication, SRM Institute of Science and Technology, for providing us with an opportunity to pursue our project under his / her mentorship. He / She provided us with the freedom and support to explore the research topics of our interest. His / Her passion for solving problems and making a difference in the world has always been inspiring.

We sincerely thank all the staff and students of Networking and Communication,, School of Computing, S.R.M Institute of Science and Technology, for their help during our project. Finally, we would like to thank our parents, family members, and friends for their unconditional love, constant support and encouragement

AKBARSAGARI MOHMMAD FAZIL - RA2111030010175

PALLAPOLU SAI VARDHAN REDDY - RA2111030010186

JAINI ESWAR - RA2111030010190

# SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
# KATTANKULATHUR – 603 203

## BONAFIDE CERTIFICATE

Certified that 18CSP107L - Minor Project report titled “**HYBRID MACHINE LEARNING MODEL TO DETECT AND MITIGATE EMAILPHISHING ATTACKS** ” is the bonafide work of “**AKBARSAGARI MOHMMAD FAZIL[RA2111030010175], PALLAPOLU SAI VARDHAN REDDY[RA2111030010186],JAINI ESWAR[RA2111030010190]”** who carried out the project work] under my supervision. Certified further, that to the best of my knowledge the work reported herein does not form any other project report or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

**SIGNATURE**
DR.S. PRABAKERAN
**SUPERVISOR**
Associate Professor,
Department of Networking
and Communications,
SRM institute of Science
and Technology

**SIGNATURE**
DR. R. ANNIE UTHRA
**PROFESSOR &HEAD**
Department Of Networking and
Communications,
SRM institute of Science
and Technology

# TABLE OF CONTENT

# ABSTRACT

The prevalence of phishing attacks has increased significantly, since this form of cyber discipline is meant to seriously test an individual as well as organizations alike. In this respect, phishing attacks are considered instances in which there is an attempt to fraudulently obtain sensitive information in a manner that is increasingly sophisticated and consequently not as easily detected by traditional methods. The project focuses on developing a use case for an advanced email classification system, incorporating some of the most high-profiled ensemble machine learning models: Support Vector Machines and XGBoost. The goal is to classify phishing and legitimate emails with high accuracy. This project will involve the use of a phishing email dataset. Text vectorization is done for preprocessing, using TF-IDF in order to capture meaningful features. It first categorizes the emails with the help of SVM having decision functions that later can be taken as input by the XGBoost model for final classification. The two-layer classification makes the detection of phishing more robust and accurate in order for high precision and recall metrics of the prediction. Additionally, a novel signature extraction and mitigation strategy against phishing has been developed. Following the extraction of phishing e-mails, the system identifies key terms and patterns indicative of phishing behavior through the TF-IDF method. All these "phishing signatures" are saved in a comma-separated variable file. This will be used for filtering incoming emails based on its similarities to known phishing patterns. This approach is lightweight, data-driven, adaptive, and therefore does not require heavy implementation or extra hardware. It is dynamic and, therefore, ever-evolving in defense against phishing attacks by constantly updating the phishing signature database. The results compare our hybrid model, SVM + XGBoost, which demonstrates higher accuracy and precision than those from traditional models like Random Forest and Logistic Regression, thus providing a more reliable and robust solution toward phishing email detection. What this suggests is that the combined use of machine learning approaches and signature-based mitigation presents a comprehensive and adaptive solution to improve cybersecurity defenses against phishing attempts.

# LIST OF FIGURES

## KEYWORDS

Phishing Detection
Machine Learning
Support Vector Machine (SVM)
XGBoost
Cybersecurity
Email Classification
Phishing Mitigation
TF-IDF
Hybrid Model
Phishing Signatures.

# 1.INTRODUCTION

## 1.1General

Phishing attacks have become one of the most pervasive forms of cybercrime, targeting individuals and organizations alike. These attacks involve tricking victims into providing sensitive information such as passwords, financial details, or personal identification data by masquerading as legitimate communications from trusted entities. With the rapid digitalization of workflows and increased reliance on email as a primary communication tool, phishing attacks have escalated, creating significant security vulnerabilities for businesses.

The need for robust detection systems to counter these attacks is more pressing than ever. Cybercriminals have adapted to conventional defences by constantly refining their tactics, making it difficult for traditional detection systems to keep up. Thus, modern cybersecurity strategies are increasingly looking towards **machine learning** as an effective tool for detecting and mitigating phishing attempts. Machine learning models, when trained on large datasets, can recognize complex patterns within phishing emails that would otherwise evade traditional detection systems.

This project introduces a **hybrid machine learning model** that combines **Support Vector Machines (SVM)** and **XGBoost** to detect phishing emails with high accuracy. The model not only focuses on identifying phishing attempts but also includes a **phishing signature extraction strategy** aimed at mitigating future attacks. By identifying and extracting unique patterns within phishing emails, the system is designed to evolve and become more effective over time, ensuring long-term protection against phishing threats.

The growing sophistication of phishing attacks necessitates advanced detection systems capable of adapting to new threats. Through the use of hybrid machine learning models and proactive mitigation strategies, this project aims to provide a cutting-edge solution that significantly enhances email security.

In this project, we propose the use of a hybrid machine learning model that combines the strengths of Support Vector Machines (SVM) and XGBoost to detect phishing emails. This system also includes a mitigation mechanism that extracts unique phishing signatures, which can be used to identify new and evolving phishing attempts. The hybrid approach allows for more accurate classification by leveraging SVM's ability to handle linear classification tasks and XGBoost's capacity to improve performance through gradient boosting. This ensures that the model is not only effective in identifying phishing attempts but also adaptable to new phishing techniques as they emerge. By implementing this advanced detection and mitigation system, organizations can significantly reduce the risk of falling victim to phishing attacks, enhancing their overall cybersecurity posture.

## 1.2 Purpose

The primary purpose of this project is to develop an **advanced phishing detection system** that leverages hybrid machine learning techniques to accurately identify and mitigate phishing attacks. Specifically, the project focuses on improving phishing detection accuracy by combining the **Support Vector Machine (SVM)** and **XGBoost** classifiers, each bringing unique strengths to the detection process.

Beyond phishing detection, the project aims to implement a phishing signature extraction strategy. This innovative approach focuses on addressing future phishing threats proactively by storing unique phishing signatures in a database. The system will not only react to phishing attempts in real-time but also create a repository of phishing patterns. This enables faster identification and mitigation of future attacks based on previously encountered phishing signatures, significantly improving response times.

Phishing attacks are becoming increasingly complex, with traditional detection systems often failing to provide adequate protection, particularly against targeted phishing tactics like spear-phishing. By leveraging machine learning, this project seeks to bridge the gap by offering a more dynamic, adaptive solution that evolves with the nature of phishing threats. The hybrid model allows for continuous learning from emerging patterns, thus enhancing its effectiveness over time.

Ultimately, the project's objective is to protect organizations from phishing attacks by strengthening their cybersecurity defences. Through this machine learning-based system, sensitive data can be safeguarded, reducing the risks of data breaches, financial loss, and reputational damage. In turn, this project contributes to building a more resilient cybersecurity framework for organizations, addressing both current and future threats.

In addition to detecting phishing emails, the project also aims to address the challenge of mitigating future phishing attacks by extracting phishing signatures. This allows the system to identify unique patterns or signatures associated with phishing emails, which can then be used to protect against similar threats in the future. By combining detection and mitigation, the project provides a comprehensive solution that not only identifies phishing attacks in real-time but also strengthens defenses against evolving phishing techniques. This dual-purpose approach ensures that organizations are better equipped to handle both known and emerging phishing threats, safeguarding sensitive information and reducing the risk of data breaches.

## 1.3Scope

The scope of this project encompasses the development and implementation of a hybrid machine learning model specifically designed to detect and mitigate phishing email attacks. It involves multiple stages, beginning with data collection and preprocessing using techniques like TF-IDF vectorization to convert email content into numerical features. This is followed by the training and implementation of two machine learning models—Support Vector Machine (SVM) and XGBoost—which are combined to form a robust detection system. The SVM model contributes by handling linear classifications, while XGBoost enhances the model's accuracy through gradient boosting, addressing non-linear patterns and boosting performance. The system not only focuses on detecting phishing emails but also includes a phishing signature extraction process. This allows the model to extract and store unique patterns found in phishing emails, which can be used for future reference to detect evolving phishing techniques.

In addition, the scope includes evaluating the performance of the hybrid model using key performance metrics such as accuracy, precision, recall, F1-score, and ROC-AUC score. This evaluation is done in comparison to other traditional machine learning models such as Random Forest and Logistic Regression, highlighting the superiority of the hybrid approach in phishing detection. Furthermore, the system is designed to be scalable, allowing for future enhancements to include the detection of other phishing vectors such as SMS, social media, and voice phishing (vishing). As a cybersecurity solution, this project also aims to provide a proactive defense mechanism by continuously updating the phishing signature database, ensuring that the model can detect new phishing patterns as they emerge. The broader scope includes the potential integration of this system into organizational email servers for real-time phishing detection and mitigation, thus fortifying an organization’s defense against email-based cyber threats.

## 1.4Objectives of the Project

The scope of this project is to develop and evaluate a hybrid machine learning model designed to detect phishing emails with a high degree of accuracy. By combining two powerful machine learning algorithms—SVM and XGBoost—we aim to create a robust system that can effectively distinguish between legitimate and phishing emails, even in the presence of imbalanced datasets. Additionally, the project incorporates a phishing signature extraction mechanism to enhance future prevention efforts, allowing the model to recognize and block new phishing attempts by identifying unique patterns in phishing emails.
The primary objectives of this project are as follows:

**1.4.1Develop a Hybrid Machine Learning Model:** We aim to create a hybrid model that combines SVM and XGBoost to detect phishing emails. SVM will serve as the base classifier to handle linear separable data, while XGBoost will be used to enhance the final classification accuracy through gradient boosting.
**1.4.2Implement Phishing Signature Extraction:** In addition to detecting phishing emails, the model will extract unique signatures or patterns from classified phishing emails. These signatures will be used to update a database of known phishing tactics, helping to mitigate future attacks.
**1.4.3Evaluate the Model's Performance:** The hybrid model will be tested on a dataset of phishing and legitimate emails, with key performance metrics such as accuracy, precision, recall, F1-score, and ROC-AUC being used to evaluate its effectiveness.
**1.4.4Compare Against Traditional Models:** To assess the benefits of the hybrid approach, we will compare its performance against traditional machine learning models such as Random Forest and Logistic Regression, which are commonly used in phishing detection tasks.
**1.4.5Propose a Mitigation Strategy:** Based on the phishing signatures extracted, we will propose a strategy to mitigate future phishing attacks by using these signatures to block similar phishing attempts.

By the end of the project, the hybrid model will not only serve as a tool for detecting phishing emails but will also contribute to a proactive defense mechanism through signature extraction. The system will continuously learn from new phishing attempts, improving its detection capabilities over time and helping to protect organizations from evolving phishing threats.

This report outlines the detailed methodology, implementation, and evaluation of the hybrid model. The results section presents a comparative analysis of the hybrid model's performance relative to traditional approaches, while the discussion highlights the advantages of the hybrid model in real-world phishing detection scenarios. Finally, the report concludes with recommendations for future work, including potential improvements to the phishing signature extraction process and the application of the model in other areas of cyberse

## 1.5Background on Phishing Attacks:

Phishing attacks have become one of the most prevalent and dangerous forms of cybercrime in recent years. As the internet and digital communication technologies continue to evolve, so too have the tactics used by malicious actors to compromise sensitive information. Phishing involves deceiving individuals into revealing confidential information, such as passwords, bank account numbers, or personal identification numbers, by masquerading as trustworthy entities. These entities can take the form of legitimate companies, government organizations, or even familiar contacts. The ultimate goal of a phishing attack is to gain unauthorized access to valuable data, which can be used for financial gain, identity theft, or other criminal activities.

Phishing attacks generally occur via email, where users receive fraudulent messages that contain links to malicious websites or attachments infected with malware. These emails often appear to be from trusted sources, tricking recipients into clicking on harmful links or providing personal information. The attackers typically craft these messages to be highly convincing, using branding, logos, and language that closely mimics legitimate communications. This high level of sophistication makes phishing a particularly insidious threat, as even vigilant users may fall victim to these schemes.

The consequences of phishing can be severe for individuals and organizations alike. In the case of individuals, phishing can lead to financial loss, identity theft, and a compromised digital footprint. For organizations, the stakes are even higher. Successful phishing attacks can result in data breaches, theft of intellectual property, loss of customer trust, and even legal liabilities. In many cases, organizations face millions of dollars in damages due to phishing-related breaches, making it one of the top cybersecurity concerns globally.

Despite significant advancements in cybersecurity defences, phishing attacks continue to proliferate due to their low cost of implementation and the relative ease with which attackers can target a wide audience. This highlights the critical need for more advanced detection and mitigation techniques to combat the growing threat. Traditional security measures, such as firewalls, antivirus software, and email filters, are no longer sufficient to protect against the increasingly sophisticated nature of phishing campaigns.

### 1.51Major Incidents in Phishing History

Several major incidents underscore the severity of phishing attacks in recent years. One of the most significant examples is the 2016 U.S. Presidential Election, where phishing played a central role in the hacking of political figures. Attackers gained access to the Democratic National Committee (DNC) email system by tricking officials into entering their login credentials on a fake Google login page. This phishing attack not only compromised sensitive information but also had far-reaching consequences for global politics.

Another high-profile incident occurred in 2013 when cybercriminals used phishing to breach Target, one of the largest U.S. retailers. Attackers gained access to Target's internal network by phishing a third-party vendor. The breach exposed the credit card information of over 40 million customers and highlighted how phishing can be used as an entry point for broader cyber-attacks, compromising sensitive data on a massive scale.

### 1.5.2Types of Phishing

Phishing attacks have diversified into various types, each exploiting different communication channels and social engineering tactics. Understanding these types is crucial for developing effective detection and mitigation strategies.

1. **Email Phishing**: This is the most common form of phishing, where attackers send fraudulent emails pretending to be from reputable sources such as banks, government institutions, or popular websites like Amazon or PayPal. These emails often contain links to fake websites designed to steal login credentials or infect users' systems with malware.
2. **Spear Phishing**: Unlike generic phishing, spear-phishing attacks target specific individuals or organizations. Attackers spend time gathering information on their targets to craft convincing messages. For example, in 2014, a spear-phishing attack compromised Sony Pictures, causing a significant data breach.
3. **Whaling**: A subtype of spear phishing, whaling targets high-profile individuals such as CEOs or government officials. These attacks often involve fraudulent emails appearing to come from trusted colleagues or partners. Since the targets hold more power and access, the damage from a successful whaling attack can be catastrophic.
4. **Vishing (Voice Phishing)**: Vishing involves attackers using phone calls to impersonate legitimate entities, such as banks or tech support, to extract sensitive information. In these scams, the attackers rely on social engineering rather than malware or infected links.
5. **Smishing (SMS Phishing)**: Smishing attacks use text messages to lure victims into providing personal information or clicking on malicious links. As mobile device usage increases globally, smishing has become a growing concern.
6. **Clone Phishing**: In clone phishing, attackers take a legitimate, previously delivered email, copy its contents, and send it again with a malicious link or attachment. Since the email appears identical to the original, recipients are more likely to trust it.
7. **Pharming**: Pharming is more technical than other phishing methods. Attackers manipulate website traffic by redirecting users from legitimate sites to fraudulent ones without their knowledge. This is typically achieved by poisoning the Domain Name System (DNS) or exploiting vulnerabilities in the user's browser.

### 1.5.3Methods Attackers Use to Evade Detection

As phishing defenses have improved, attackers have developed increasingly sophisticated methods to bypass detection systems. Some of the key techniques used by phishing attackers include:

1. **Spoofing**: One of the most common methods, spoofing involves forging email headers, addresses, or domains to make phishing emails appear as though they are coming from legitimate sources. For example, attackers may use email addresses that closely resemble those of legitimate organizations by changing one or two characters (e.g., "g00gle.com" instead of "google.com").
2. **Social Engineering**: Phishing relies heavily on social engineering, a technique in which attackers manipulate human behavior to trick users into disclosing sensitive information. This often involves creating a sense of urgency, fear, or trust to encourage users to click on malicious links or download dangerous attachments.
3. **Malicious Attachments and Links**: Phishing emails often include attachments or links that, when opened, download malware onto the victim's device or direct them to a fake website designed to steal login credentials.
4. **Use of HTTPS and SSL Certificates**: Attackers increasingly use SSL certificates and HTTPS to make their phishing sites appear more legitimate. Since users are often taught to trust websites with "https://" in the URL, this can easily mislead them into believing they are on a legitimate site.
5. **URL Shorteners**: Phishers often use URL shorteners (e.g., bit.ly) to hide the true destination of a malicious link, making it more difficult for users and automated systems to identify phishing attempts.
6. **Targeted Payloads**: Sophisticated phishing attacks can deliver custom payloads to the victim's device, tailored to the operating system or browser in use, which helps evade standard detection methods.

### 1.5.4.Statistics and Real-World Examples of Damage

The damage caused by phishing attacks can be enormous, both financially and in terms of reputation. According to the FBI's Internet Crime Complaint Center (IC3), phishing was the most reported type of cybercrime in 2020, with over 241,000 incidents reported. The total financial losses from phishing attacks in the same year exceeded $4.2 billion globally. These numbers highlight the need for more advanced detection and mitigation strategies.

For example, in 2018, phishing was responsible for the largest data breach in history at Facebook and Google, which saw cybercriminals successfully steal over $100 million through a fake invoice scam. Attackers impersonated a Taiwanese hardware manufacturer and sent fraudulent invoices to the tech giants, which were paid without raising suspicion.

In another instance, the UK's National Health Service (NHS) was heavily impacted by a phishing-related ransomware attack in 2017. Although the phishing email was the initial entry point, the attack ultimately affected hospital systems, causing widespread disruption to healthcare services across the country.

These real-world examples and statistics underscore the critical need for ongoing research and development of advanced phishing detection systems, such as the one proposed in this project, to combat the evolving threat landscape.

## 1.6.Importance of Machine Learning in Detecting Phishing:

The emergence of machine learning (ML) has brought new opportunities to address cybersecurity challenges, particularly in the detection and prevention of phishing attacks. Machine learning algorithms are designed to learn from data, recognize patterns, and make predictions, which makes them well-suited for detecting phishing emails that often share similar characteristics. In contrast to rule-based systems that require predefined patterns to identify phishing emails, machine learning models can continuously improve by learning from new data, making them more adaptable to the constantly evolving nature of phishing tactics.

Machine learning offers several key advantages in the fight against phishing. First, it can automate the detection process, significantly reducing the time required to identify and respond to phishing attempts. This is particularly important in large organizations that receive a high volume of emails daily, where manual review would be impractical. Second, machine learning models can analyze vast amounts of data, identifying subtle patterns and correlations that might be missed by traditional rule-based systems. These patterns can include suspicious sender domains, unusual content structures, or anomalous behaviors in links and attachments.

Another critical advantage of machine learning in phishing detection is its ability to handle imbalanced datasets. In real-world applications, phishing emails typically make up a small percentage of overall email traffic, creating a challenge for traditional algorithms that may struggle to detect minority classes effectively. Machine learning models, particularly those designed for imbalanced data, such as Support Vector Machines (SVM) and XGBoost, are capable of identifying phishing emails even when they are rare, ensuring higher detection accuracy.

Moreover, machine learning models can evolve and adapt as phishing strategies change. Attackers are constantly devising new ways to bypass traditional defenses, making static security systems obsolete over time. Machine learning algorithms can be retrained on new data, allowing them to stay ahead of emerging threats. This adaptability is crucial for maintaining robust defenses in the face of dynamic and ever-changing phishing campaigns.

In recent years, the application of hybrid machine learning models has gained traction as a more effective approach to phishing detection. Hybrid models combine the strengths of multiple machine learning techniques, enabling more accurate and reliable detection. In this project, we propose a hybrid model that integrates Support Vector Machines (SVM) with XGBoost, aiming to enhance phishing detection accuracy while mitigating future attacks through phishing signature extraction.

### 1.6.1.Advantages of Machine Learning in Handling Large and Dynamic Datasets

One of the key reasons machine learning is so effective in phishing detection is its capacity to handle large, dynamic datasets. In contrast to traditional detection systems, which rely on manually curated lists of phishing URLs or email characteristics, machine learning algorithms can process vast amounts of data in real-time. This scalability is crucial in the modern digital landscape, where billions of emails are sent daily, and phishing campaigns can evolve rapidly.

Machine learning models, particularly those based on supervised learning, can be trained on historical data that contains both legitimate and phishing emails. By learning from these examples, the model can make informed predictions about new, unseen emails based on patterns and features it has previously encountered. For example, a trained model can analyze the content, sender information, and embedded links in an email to determine whether it is likely to be a phishing attempt. This level of automation allows organizations to stay ahead of attackers by flagging suspicious emails even as new phishing techniques emerge.

Furthermore, machine learning models can continuously update and improve as they encounter new data. This is especially beneficial for phishing detection, where attackers frequently adapt their methods to bypass traditional filters. By leveraging large, diverse datasets, machine learning models can identify subtle, evolving characteristics in phishing emails that would be impossible to detect through static rule-based systems.

### 1.6.2.Capturing Complex Patterns and Relationships

Machine learning's ability to capture complex patterns and relationships within data sets it apart from traditional approaches. Phishing attacks are often designed to deceive users and bypass simple filters by using variations in language, structure, and behavior. Traditional systems, which rely on predefined rules and signatures, struggle to keep up with these dynamic changes. However, machine learning algorithms, especially those using techniques like natural language processing (NLP), can detect nuanced features in emails, such as suspicious wording, sentence structure, or even stylistic differences that may indicate a phishing attempt.

For example, an ML model can learn to differentiate between legitimate corporate emails and phishing attempts by recognizing subtle linguistic patterns, such as urgency in the subject line ("immediate action required") or the use of informal greetings in otherwise professional settings. Such patterns, when observed in isolation, may not trigger traditional detection mechanisms, but machine learning algorithms can analyze them in conjunction with other factors like sender reputation, domain age, and the presence of malicious links.

Moreover, ML models excel at detecting relationships between variables that may not be immediately apparent. Phishing emails often include a combination of deceptive techniques—such as spoofed domains, hidden links, or attachments—that, when considered together, form a unique signature of an attack. Machine learning can model these interdependencies, allowing the system to identify phishing attempts with higher accuracy compared to methods that focus on individual factors in isolation.

### 1.6.3.Real-World Applications of Machine Learning in Phishing Detection

The practical application of machine learning in phishing detection is already being realized by various cybersecurity solutions and email security platforms. Many of these systems employ supervised learning models trained on large datasets of known phishing and legitimate emails, using features such as email content, metadata, and behavior patterns.

For instance, Google's Gmail employs machine learning to detect phishing emails, filtering over 100 million phishing emails daily. By analyzing historical data from billions of emails, the system can effectively flag malicious messages before they reach users' inboxes. Gmail's phishing detection system is based on a blend of machine learning algorithms, including neural networks, that continually adapt to new attack vectors as phishing tactics evolve.

Similarly, Microsoft's Office 365 Advanced Threat Protection also leverages machine learning to protect against phishing and other email-based threats. The platform uses ML models to assess the likelihood that an email is part of a phishing campaign, considering a wide range of factors such as sender reputation, message content, and the presence of anomalies that suggest fraudulent activity. By incorporating machine learning, Office 365 can detect and block phishing emails with greater precision, even if the specific phishing techniques have not been encountered before.

In the academic field, multiple studies have demonstrated the efficacy of machine learning in phishing detection. One prominent example is a research paper by Rao and Ali, who developed a machine learning-based phishing detection system that achieved a detection accuracy of over 95% using a combination of NLP and supervised learning models. Their system was able to detect phishing emails by analyzing features such as word frequency, hyperlink behavior, and metadata.

### 1.6.4.Hybrid Models as a Solution

To address some of these limitations, hybrid machine learning models have been developed that combine the strengths of different algorithms. For example, integrating a Support Vector Machine (SVM) with an XGBoost classifier can improve phishing detection by leveraging the strengths of both models. SVM is well-suited for handling high-dimensional data and is particularly effective in binary classification tasks, such as determining whether an email is phishing or legitimate. On the other hand, XGBoost, a powerful gradient-boosting algorithm, excels in processing large datasets and identifying complex, nonlinear relationships between features.

By combining these models in a hybrid system, it is possible to achieve higher accuracy and robustness in phishing detection. The SVM can act as an initial filter, quickly classifying emails based on linear features, while the XGBoost model can further analyze emails flagged as potentially suspicious, examining more complex patterns to confirm or refute the initial classification.

## 1.7.Overview of Hybrid Machine Learning Models

Hybrid machine learning models represent a significant advancement in the field of predictive analytics, particularly in complex problem domains like phishing detection. These models combine two or more individual machine learning algorithms to form an ensemble, resulting in a more robust and accurate prediction system. The primary motivation behind hybrid machine learning models is to leverage the strengths of multiple algorithms while mitigating their individual weaknesses. In phishing detection, this approach is especially useful due to the wide variety of phishing strategies employed by attackers and the diverse characteristics of phishing emails.

### 1.7.1.Defining Hybrid Machine Learning Models and Model Ensembling

A hybrid machine learning model is essentially a composite model that combines the predictions or outputs of multiple base models to improve overall performance. This concept is closely related to model ensembling, a machine learning technique where different models are trained on the same task, and their predictions are combined to produce a final output. There are several methods of ensembling, such as bagging, boosting, and stacking. These approaches aim to reduce overfitting, improve generalization, and enhance the overall accuracy of the prediction system.

In phishing detection, the complexity of the problem lies in the wide variation in phishing strategies, making it difficult for a single model to consistently perform well across all attack types. While one model might excel at detecting certain patterns in email content, it may struggle with identifying more subtle characteristics found in other phishing attempts. A hybrid model allows for a more comprehensive analysis, improving detection capabilities by aggregating the insights from multiple algorithms.

In the context of our project, the hybrid model combines a Support Vector Machine (SVM) and XGBoost classifier. These two algorithms complement each other in various ways, allowing the hybrid model to outperform individual models when detecting phishing attacks.

### 1.7.2.Advantages of Hybrid Models Over Single Models

The main advantage of hybrid models is their ability to capitalize on the strengths of multiple algorithms while compensating for each model's weaknesses. Single models are often limited by their inherent biases, meaning that they may perform well on specific types of data but poorly on others. By using a hybrid model, this limitation can be mitigated, as each algorithm brings different strengths to the table.

For example, SVM is highly effective at handling high-dimensional data and is particularly suitable for binary classification tasks, which makes it a strong candidate for the phishing detection problem. SVM is known for creating clear decision boundaries between phishing and legitimate emails based on features such as email structure, content, and metadata. However, it may struggle when it comes to capturing complex, nonlinear relationships in the data.

On the other hand, XGBoost (Extreme Gradient Boosting) is a powerful algorithm known for its ability to handle large datasets and complex, nonlinear relationships between variables. XGBoost works by creating an ensemble of decision trees, where each tree improves on the mistakes of the previous one. This makes XGBoost particularly well-suited for phishing detection, where attackers often use subtle, nonlinear techniques to evade detection.

By combining the SVM and XGBoost models, we can address their individual limitations. The SVM excels at drawing clear boundaries between phishing and legitimate emails, while XGBoost is adept at uncovering hidden relationships in the data that may not be obvious at first glance. Together, they form a more comprehensive detection system that is both precise and adaptable to different types of phishing attacks.

### 1.7.3.How SVM and XGBoost Work Together in Phishing Detection

In our hybrid model, SVM and XGBoost work together by first applying SVM to the dataset to classify emails as either phishing or legitimate based on linear features such as the presence of certain keywords, email format, and sender information. SVM provides a strong initial filter for more obvious phishing attacks, drawing on its strength in binary classification.

Once SVM has performed its initial classification, the XGBoost model further analyzes emails that are flagged as potentially suspicious by SVM. This second step allows the system to capture more subtle, nonlinear patterns that SVM might have missed. For instance, XGBoost can identify sophisticated phishing attacks that rely on social engineering techniques, as well as attacks where the visual appearance of the email mimics a trusted source.

The combination of these two models leads to a higher detection accuracy, as each model compensates for the other's weaknesses. By using SVM as the first line of defense and XGBoost as a more detailed secondary check, the hybrid model ensures that even highly sophisticated phishing attempts are accurately detected. Furthermore, this approach helps to reduce false positives, as the XGBoost model provides an additional layer of verification for emails that may have been incorrectly flagged by the SVM model.

In summary, the hybrid model significantly enhances phishing detection by leveraging the strengths of both SVM and XGBoost. Together, these models offer a more comprehensive and adaptable solution to detecting phishing attacks in a constantly evolving cybersecurity landscape.

## 1.8.Contributions:

This project introduces several key contributions to the cybersecurity field, especially in phishing detection and mitigation. Firstly, it presents a **hybrid machine learning model** that combines Support Vector Machines (SVM) and XGBoost classifiers. This hybrid approach leverages SVM’s ability to efficiently handle linear classification tasks and XGBoost’s boosting techniques, resulting in enhanced detection accuracy, precision, and recall. By improving upon traditional models such as Random Forest and Logistic Regression, the project sets a new standard for phishing email detection.

Secondly, the project implements an innovative **phishing signature extraction strategy**, which not only detects phishing attempts but also captures and stores unique phishing signatures. These signatures can be referenced in the future, allowing the system to quickly recognize patterns and respond to new phishing threats more effectively. This approach contributes to both proactive cybersecurity defense and enhanced phishing mitigation.

## 1.9.Practical Implications

From a practical standpoint, this project can **significantly enhance organizational security** by reducing the risk of data breaches, financial losses, and reputational damage caused by phishing attacks. The hybrid model's increased accuracy ensures that phishing emails are more effectively detected, safeguarding sensitive data. Additionally, the phishing signature extraction feature promotes **resource efficiency**, as previously identified patterns can be quickly compared, minimizing the need to reprocess data and saving computational effort.

The system is also designed to be **scalable and adaptable**, continuously learning from new phishing techniques and signatures, which makes it applicable across various industries, from small businesses to large enterprises. Lastly, the research behind this project contributes to the broader **cybersecurity community**, offering a foundation for future advancements in email security and phishing attack mitigation strategies.

# CHAPTER 2

# LITERATURE REVIEW

## 2.1Research Findings

I. Champa, M. F. Rabbi and M. F. Zibran, "Curated Datasets and Feature Analysis forPhishing Email Detection with Machine Learning," (ICMI), Mt Pleasant, MI, USA, 2024.This paper highlights the urgent need for effective detection methods.The authors address the challenge of scarce, well-curated datasets for phishing detection, leading to the creation of seven publicly available datasets that are ready for machine learning applications. The study also investigates the features of emails that are most influential in distinguishing phishing from legitimate emails, applying five machine learning algorithms to derive insights.The findings are constrained to English emails, which may limit their applicability to other languages and contexts .The analysis did not consider the structure of URLs or email attachments, which could provide deeper insights into phishing attempts.

A. Chien and P. Khethavath, "Email Feature Classification and Analysis of Phishing Email Detection Using Machine Learning Techniques," (CSDE), Nadi, Fiji, 2023.The study analyzes 16,906 emails using various machine learning techniques to improve detection.Two experiments were conducted: one to classify legitimate advertisements versus phishing emails, and another to distinguish between legitimate and phishing emails. Results indicated that while phishing emails could be identified, distinguishing between advertisements and phishing was challenging due to overlapping features.The research faced significant challenges due to the unavailability of diverse real email datasets.

1st Tosin Ige dept. of Computer Science The University of Texas The paper discusses the vulnerabilities in modern systems that can be exploited by cyberattacks, emphasizing the importance of early detection and prevention methods.It highlights the effectiveness of machine learning in cybersecurity, particularly in detecting various types of attacks, including phishing and malware, while noting the underperformance of Naïve Bayes classifiers in certain tasks.The section categorizes current phishing detection models, focusing on Bayesian-based classifiers like Naïve Bayes, which struggle due to their strong independence assumptions.It also reviews non-Bayesian classifiers such as Decision Trees and Random Forests, which generally outperform Naïve Bayes in various classification tasks.The findings underscore the necessity for further exploration of machine learning techniques for detecting drive-by downloads and improving Naïve Bayes classifiers. It also emphasizes the need for advancements in SQL Injection detection methods to address the limitations of existing approaches in identifying compromised databases.

I. Champa, F. Rabbi and M. F. Zibran, "Why Phishing Emails Escape Detection: A Closer Look at the Failure Points," 2024. This study addresses the challenges in detecting phishing emails, particularly the inadequacy of existing datasets for machine learning applications, and introduces 11 curated datasets for research use. The findings highlight how scammers craft emails that closely mimic legitimate communication, complicating detection efforts.

The study primarily utilized five well-known machine learning algorithms, excluding deep learning models, which may limit the exploration of more advanced detection techniques.

The unjan and R. Prasad, "Phishing Email Detection Using Machine Learning: A Critical Review," 2024. The study emphasizes the challenges of automatic phishing email detection and the potential of machine learning (ML) algorithms like SVM, Naive Bayes (NB), and LSTM in improving detection accuracy. The research aims to integrate natural language processing (NLP) with ML techniques to enhance the identification of phishing emails, providing insights into current methodologies and datasets. Traditional machine learning techniques require manual feature engineering, complicating their application in dynamic data environments.

1st Takeshi Matsuda dept. Management and Information Technology Hannan University The paper proposes a novel email classification method that utilizes modality representations and dimensionality reduction to enhance spam detection, particularly for evolving scam emails. It emphasizes the need for continuous updates to the corpus of scam emails and the polarity dictionary used for sentiment analysis, given the changing nature of fraudulent techniques. The study introduces ten modalities related to mental attitudes and applies unsupervised learning to classify scam emails from legitimate ones. The effectiveness of the proposed method is validated through 3D visualizations using UMAP, demonstrating its capability to track changes in scam email structures over time. The research highlights the need for ongoing development of advanced detection methods that can adapt to the continuously changing landscape of email scams. Future work will aim to enhance the classification process, moving beyond simple binary determinations of email legitimacy.

1st S. PriyaDept. of Information Technology Manipal Institute of Technology The paper discusses the rise of phishing attacks, which target sensitive user information through fraudulent links, highlighting the increasing sophistication of hackers. Phishing attacks account for a significant portion of cybersecurity threats, with a focus on the methods used to deceive users into revealing personal information. Machine learning techniques, including Decision Trees and Random Forests, are employed to classify phishing websites, with Random Forests effectively managing overfitting. Support Vector Machines also show promise, but selecting the right kernel function can be complex. The paper concludes that current phishing detection methods may not suffice against evolving attack strategies, necessitating the development of adaptive techniques. Insights into existing tools and their limitations can guide future research efforts to create more effective solutions for combating phishing attacks.

## 2.2.Current Approaches to Phishing Detection

Phishing detection has evolved significantly over the past few decades, with machine learning playing a central role in the development of more sophisticated and effective detection systems. Various machine learning models, including supervised, unsupervised, and deep learning techniques, have been explored in the quest to improve the identification of phishing emails.

Supervised Learning Approaches: Supervised learning models have been widely adopted in phishing detection, where labeled datasets of phishing and legitimate emails are used to train classifiers. Popular algorithms in this category include Logistic Regression, Naïve Bayes, Decision Trees, Random Forests, and Support Vector Machines (SVM). These models work by learning patterns in the training data, such as specific keywords, email structure, or malicious links, that differentiate phishing emails from legitimate ones. For example, Random Forest classifiers have been shown to perform well when detecting phishing emails by analyzing textual and metadata features such as email headers, URLs, and domain characteristics. However, their performance can be constrained by the diversity and quality of the training data.

Unsupervised Learning Approaches: Unsupervised learning models, unlike supervised models, do not rely on labeled datasets. Instead, they try to detect anomalies or unusual patterns in the data that deviate from normal email behaviors. Common unsupervised techniques used in phishing detection include Clustering Algorithms (such as K-Means or DBSCAN) and Anomaly Detection methods. These approaches can be particularly useful when dealing with novel phishing attacks that were not part of the original training dataset, enabling systems to identify suspicious emails based on behavioral changes or outlier detection. However, the lack of labeled data poses challenges in achieving high accuracy, and these methods often generate a higher number of false positives.

Deep Learning Approaches: In recent years, Deep Learning techniques such as Artificial Neural Networks (ANNs), Convolutional Neural Networks (CNNs), and Recurrent Neural Networks (RNNs) have been increasingly applied to phishing detection. These models can process large amounts of unstructured data, making them suitable for analyzing email content and extracting features directly from raw data. CNNs, for instance, have been employed to detect phishing websites by examining visual similarities to legitimate sites. RNNs and Long Short-Term Memory (LSTM) networks have been used to process the sequence of words in an email or URL, capturing context and dependencies that are crucial for detecting phishing content. While deep learning models are powerful and capable of identifying complex patterns, they come with higher computational costs and require substantial labeled data for effective training.

## 2.3.Limitations of Standalone Models

Standalone machine learning models, while effective in some cases, face several challenges when used in isolation for phishing detection. One of the primary limitations is that these models tend to specialize in certain types of phishing attacks and may struggle to generalize across diverse and evolving phishing techniques. The evolving nature of phishing, where attackers continuously adapt their methods to bypass detection, presents a significant obstacle for standalone models.

For instance, Support Vector Machines (SVM) are known for their effectiveness in binary classification problems but can struggle with the nonlinear relationships and complex patterns seen in some phishing emails. Similarly, Decision Trees may easily overfit on a small or biased training dataset, leading to poor generalization when encountering new types of phishing attacks. Models like Logistic Regression and Naïve Bayes often rely on simpler feature sets, which can limit their ability to capture more sophisticated phishing tactics, such as highly targeted spear-phishing attacks or emails that use advanced social engineering techniques.

Another limitation is the issue of overfitting, where models perform well on training data but fail to generalize to unseen data. This is particularly problematic in phishing detection, where new attack variants frequently emerge. Standalone models may also produce a high number of false positives, leading to legitimate emails being misclassified as phishing attempts. This can cause disruptions in normal business operations and erode trust in the detection system.

In addition, phishing attacks are often subtle, relying on a combination of social engineering and technical deception. Standalone models that rely on basic email features such as URLs, subject lines, or sender information may miss these nuanced indicators, especially when attackers use techniques like domain spoofing, homograph attacks, or visual impersonation to make their phishing emails appear legitimate.

## 2.4.Hybrid Approaches

To overcome the limitations of standalone models, researchers and cybersecurity professionals have increasingly turned to hybrid machine learning models. Hybrid models combine two or more machine learning algorithms, often leveraging the strengths of each while compensating for their weaknesses. By using multiple models in tandem, hybrid approaches can significantly improve phishing detection accuracy and adaptability, especially when dealing with new and evolving phishing techniques.

One of the most common strategies in hybrid models is ensemble learning, where different models are trained on the same data and their predictions are combined to produce a final classification. Ensemble methods, such as bagging and boosting, are particularly effective in reducing overfitting and improving the generalization of the model. Boosting algorithms like XGBoost are designed to sequentially correct the errors of previous models, allowing the ensemble to achieve high accuracy on challenging datasets.

In phishing detection, hybrid models often combine a simple, fast model (such as Logistic Regression or Naïve Bayes) with more complex models (such as SVM, Random Forest, or XGBoost) to achieve a balance between speed and accuracy. For instance, a hybrid model might first use SVM to quickly classify the majority of emails and then apply XGBoost to analyze the more complex cases where the decision is less clear. This two-step approach allows for a more efficient processing pipeline, reducing computational overhead while maintaining high detection rates.

Hybrid models also offer better resilience to adversarial attacks, where attackers deliberately craft phishing emails designed to evade detection by machine learning models. By combining multiple models that analyze different aspects of an email, hybrid systems are less likely to be tricked by attackers who exploit the weaknesses of a single model. For example, an attacker might design a phishing email that bypasses a traditional model by altering certain keywords or structures. However, a hybrid model that uses SVM to check the structural integrity of an email and XGBoost to analyze its content would be more likely to detect the phishing attempt.

In our project, we have employed a hybrid model that combines Support Vector Machine (SVM) and XGBoost to address the diverse challenges posed by phishing attacks. SVM excels at classifying linear and well-structured data, making it ideal for detecting certain types of phishing emails based on predefined patterns. XGBoost, on the other hand, is a powerful gradient boosting algorithm that can handle more complex, nonlinear relationships in the data, capturing sophisticated phishing techniques that may bypass simpler models. By integrating these two models, our hybrid system achieves higher accuracy and adaptability than either model could individually.

Hybrid models are also more adaptable to new phishing strategies. Since phishing techniques evolve rapidly, a static detection model quickly becomes outdated. In contrast, a hybrid model can incorporate online learning mechanisms, allowing the system to update itself based on new data, thereby staying relevant as new attack patterns emerge.

In conclusion, while standalone machine learning models have made significant strides in phishing detection, hybrid approaches offer a more comprehensive and adaptable solution to combating phishing attacks. By combining the strengths of multiple models, hybrid systems can detect a wider range of phishing techniques, improve detection accuracy, and reduce false positives, making them an essential tool for organizations looking to bolster their cybersecurity defenses against phishing attacks.

## 3. EXISTING METHOD

Traditional solutions for phishing detection rely on rule-based systems and signature-based approaches. All these techniques involve developing some rules or predefined patterns to spot phishing emails through blacklisting known malicious URLs, determining suspicious phrases, or detecting formatting and language features common in phishing attacks. The utility of these methods lies in recognizing the already-identified tactics of phishing; however, critical limitations found in them make the methods not that effective to fight against the threats that change with each passing day.

These Rule-Based Systems work by exercising a fixed set of predefined rules to detect emails containing particular keywords, phrases, or suspicious patterns. However, unfortunately due to this rigidity, most of these systems cannot adapt to new and advanced phishing attempts. It is quite easier for phishers to dodge these rules by making small changes to the contents or format of their emails, making rule-based systems pretty less credible in offering feasible detection. This rigidity essentially forms a foundation for the development of more sensitive detection methods, as the cybercrime ways change at an incredible pace.

The signature-based detection matches the emails coming in with a database of known phishing signatures, meaning those particular patterns or traits common in phishing emails. Although this technology works well in detecting established threats, it has no method to trace zero-day phishing attacks or those with strategies yet to make it into the database. Thus, the comprehensive protection of signature-based detection against upcoming phishing threats is restricted.

Information Filtering can be attributed to content filtering, which inspects the text, links, and attachments of an email in order to find hazardous information. This method usually tends to have a very high level of false positives because there will be some specific keywords or styles of formatting that would trigger the filters. Also, this quite often undermines users' trust in the mail system and drives down the effectiveness of all security measures.

Heuristic-based approaches leverage heuristics applied to the behavioral and characteristic features of emails to identify phishing attempts. These methods give better flexibility than the classic rule-based mechanisms; however, they still cannot catch up with the dynamism and evolution in phishing attacks, thus easily facing problems of accuracy and a rise in false positives.

# 4.PROPOSED METHODOLOGY

## 4.1Introduction to the Hybrid Machine Learning Approach

The hybrid machine learning approach proposed in this project is designed to address the growing sophistication of phishing attacks. Traditional detection methods, such as rule-based systems or standalone machine learning models, often struggle to keep up with the evolving tactics employed by attackers. These methods may fail to capture the subtle differences between legitimate and phishing emails, resulting in a high number of false positives or missed phishing attempts.

By integrating both Support Vector Machine (SVM) and XGBoost classifiers, the proposed hybrid model leverages the distinct advantages of each algorithm. SVM is particularly effective at establishing a clear decision boundary between phishing and legitimate emails, utilizing features extracted from email content and metadata. These features include elements like email text, header information, and metadata that distinguish phishing attempts from legitimate communications.

XGBoost, on the other hand, enhances the model's ability to focus on hard-to-detect phishing emails. By emphasizing misclassified samples, XGBoost improves detection accuracy, making it highly effective at detecting sophisticated and targeted phishing attacks, such as spear-phishing and other advanced techniques. This complementary relationship between SVM and XGBoost ensures that the hybrid model can detect both simple and complex phishing attempts, resulting in a more robust and reliable phishing detection system.

Additionally, the hybrid model is designed to adapt over time, continuously learning from new phishing attempts and refining its detection capabilities. This adaptability is crucial in a field where phishing strategies are constantly evolving. The combined strengths of SVM and XGBoost, along with the model’s adaptive nature, offer a powerful solution to the challenge of detecting and mitigating phishing attacks, safeguarding organizations from data breaches and other malicious activities.

The suggested method overcomes the limitations of standard phishing detection strategies by incorporating a hybrid machine learning (ML) framework that leverages the strengths of Support Vector Machine (SVM) and XGBoost classifiers. This strategy is intended to improve the accuracy of phishing detection, reduce false positives, and adapt to new phishing techniques.

Hybrid ML Models:
The Support Vector Machine (SVM) classifier serves as the model's first stage. It establishes a strong decision border by increasing the gap between phishing and legitimate emails, providing a clear distinction between the two classes. SVM excels at processing

high-dimensional data, making it ideal for extracting complicated features from email content and metadata.

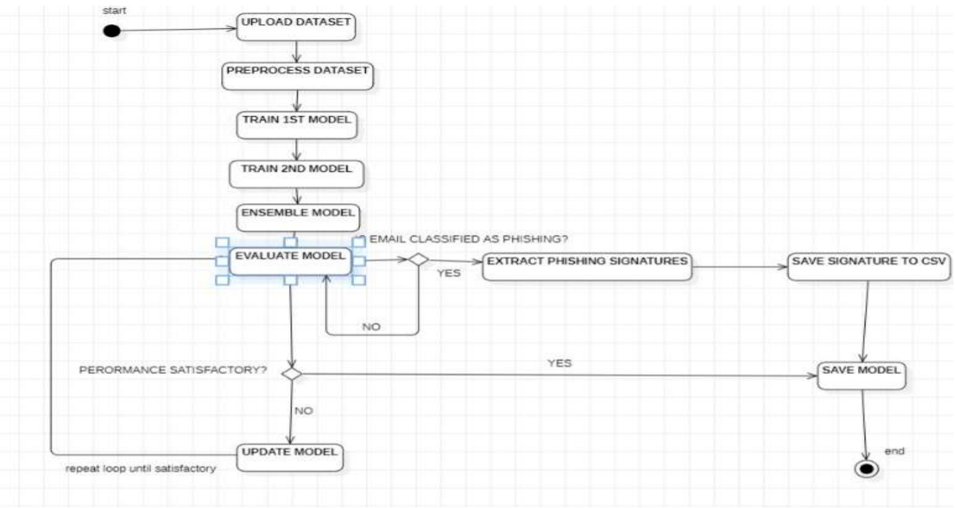
XGBoost: The SVM model's outputs are fed into the XGBoost classifier, which focusses on improving classification, particularly for difficult-to-detect instances. XGBoost, noted for its gradient boosting capabilities, iteratively improves the model by focusing on misclassified samples, increasing overall detection accuracy and lowering the risk of false positives.

Adaptive Framework: The proposed solution is meant to be adaptable, so it can change when new phishing tactics emerge. The solution maintains its effectiveness against developing threats by regularly updating the phishing signature database and retraining the ML model with new data.

The combination of SVM and XGBoost ensures that the model can deal with both simple and complex phishing attempts, making it more robust and trustworthy than older techniques.

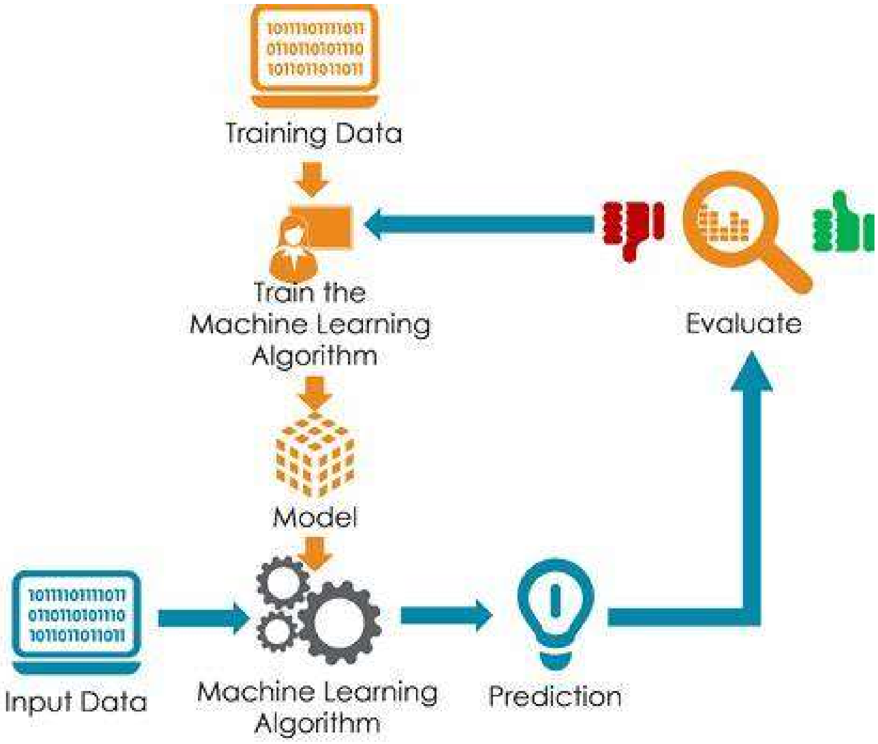
Scalability and implementation:
The entire framework is lightweight and can be built with common software tools like Python. It is compatible with a variety of contexts, including email servers and cloud-based security solutions, and does not require any specialized hardware. The system's modular design enables quick updates and integration with existing security infrastructures, making it scalable and viable for real-world applications.

This suggested method provides a complete, data-driven approach to phishing detection and mitigation, addressing the drawbacks of existing methods through the use of advanced machine learning algorithms and novel phishing signature extraction strategies.



Workflow of our hybrid system of phishing detection and mitigation. Starting with uploading the dataset, pre-processing, training of two separate models are involved. The two models will be then combined in one single ensemble method so that accuracy of classification improves. After the model is trained, it is evaluated for its ability to detect phishing emails. If the email is classified as phishing, then it will extract the phishing signature and store it in a CSV file for further references. In case the result of the model performance is good, the process will end with saving the model. Otherwise, the model goes to further updating and training to update until desired accuracy can be achieved. The performance enhancement and adaptation to newer phishing techniques go on in an iterative loop, hence developing a robust database of phishing signatures for future detection and mitigation.

**Architecture diagram for proposed model:**



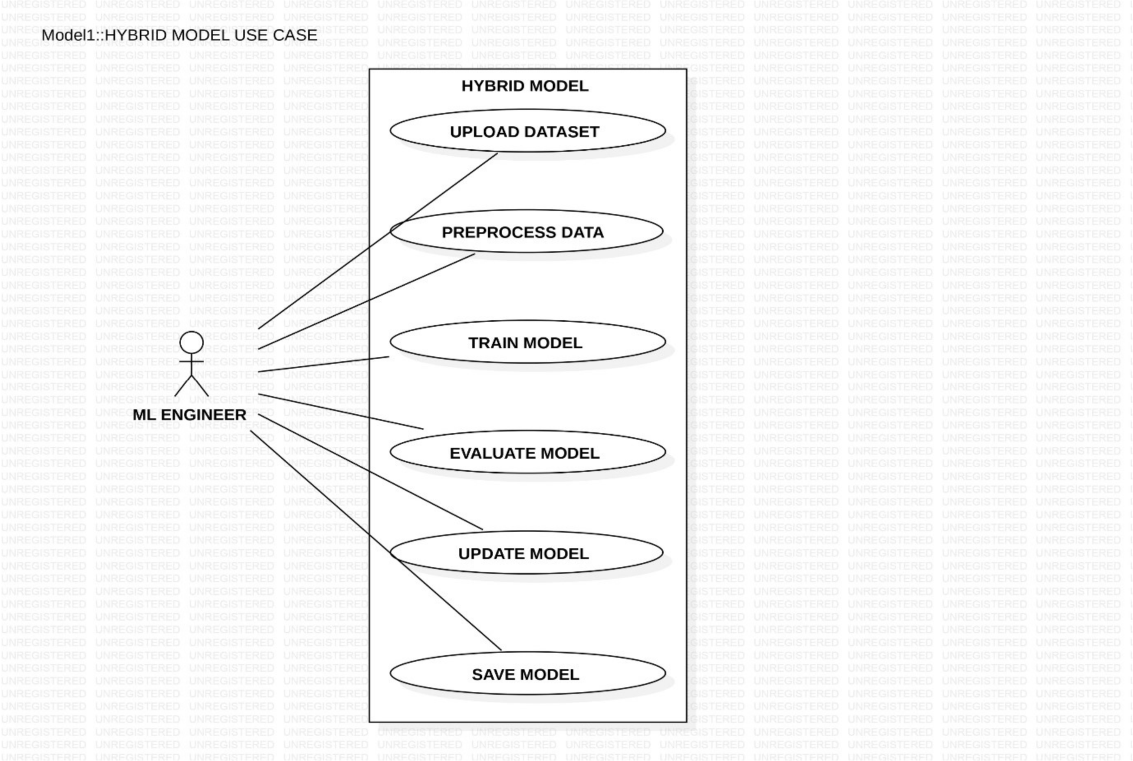
Training Data
Train the Machine Learning Algorithm
Evaluate
Model
Input Data
Machine Learning Algorithm
Prediction

Dataset : Features in this dataset include email text and email type. This numeric target variable can be described as 1-Phishing email and 0-Legitimate email. Preprocessing for email text in the dataset involved cleaning, tokenizing, and vectorizing using TF-IDF.

The proposed methodology is a robust, data-driven solution to the detection and mitigation of phishing attacks through the integration of advanced machine learning techniques with practical phishing signature extraction. First, cleaning of the dataset of emails was performed, with data preprocessing labeled as either phishing or non-phishing by getting rid of the irrelevant data and taking care of the missing values. The model will be trained based on high-quality input. Next, TF-IDF normalizes the contents and turns your text into numerical features suitable for machine learning algorithms. That is, such a representation will help the model understand the very essence of the use of words in the context of the email; therefore, it will establish a difference between phishing and legitimate mail.

The next step is SVM model training, where the SVM takes the TF-IDF features for classifying the emails. It is at this step in the process that the model provides decision scores regarding how far apart they are from their hyperplane—a theoretical boundary that separates two classes. These scores are fed into the next model, an XGBoost classifier, for refinement. XGBoost learns off those decision scores from its gradient-boosting algorithm to build a number of decision trees that better classify and achieve accuracy. Thus, this two-step approach leverages the strengths of both models and improves the performance of phishing email detection in general.

**Use case diagram for proposed model**



Moreover, to complement the classification process, a phishing signature extraction strategy has been employed to further the mitigation efforts. These are filtered out of the dataset, narrowing the focus onto confirmed phishing emails only. The most important terms from these phishing emails are then extracted using TF-IDF vectorization. The mechanism of TF-IDF ensures that unique words—that happen infrequently in legitimate emails but surface in phishing attempts—are highlighted. The resulting top terms will be compiled into a "signature" list that captures the distinctive characteristics of phishing emails. This list of signatures is then written to a CSV file, phishing_signatures.csv, which can be used as a database for phishing signatures in future filtering rules. Integration of these extracted signatures into filtering makes email filtering systems more efficient and proactive against phishing attacks.

### 4.1Comparison with Traditional models:

The hybrid model is outperforming all the other models studied herein, in comparison with Random Forest and Logistic Regression regarding accuracy and recall. While a Random Forest model does not easily overfit due to its ensemble structure, results deteriorate for high-dimensional data, slightly below the considered metrics, with an accuracy of 96% and a precision of 94%. In contrast, despite the simplicity and interpretability of Logistic Regression, it lacks the ability to learn complicated relationships within the text data and achieved an accuracy of 96% with 94% precision. Overall, our hybrid model gives the best results, especially when dealing with imbalanced and noisy datasets, thus making it more reliable for phishing detection tasks on metrics of 97% accuracy and 95% precision.

### 4.2Pseudo Code:

**1.Data Preprocessing**

```
BEGIN
   Load email dataset (phishing and legitimate emails)
   Clean the dataset by removing missing values, stop words, punctuations, and special characters
   Apply TF-IDF vectorization to convert email text into numerical features
   Split dataset into training set (80%) and testing set (20%)
END
```

**2.Train the SVM model**

```
BEGIN
   Train SVM model using the TF-IDF features from the training set
   For each email in the training set:
      Compute the decision function (classification score) of the SVM model
   Output SVM decision function results as input features for the XGBoost model
END
```

**3.Train the XGBoost Model**

```
BEGIN
   Train XGBoost model using SVM decision function outputs as input
   Optimize the XGBoost model to improve accuracy and precision
   Perform cross-validation to avoid overfitting
END
```

**4.Evaluate the Hybrid Model(SVM+XGBOOST)**

```
BEGIN
   Test the hybrid model on the testing set
   Calculate the following metrics:
      Accuracy, Precision, Recall, F1-Score, and ROC-AUC
   Compare the hybrid model's results to traditional models (Random Forest, Logistic Regression)
   Print and record the evaluation metrics for all models
END
```

**5.Phising signature extraction (Mitigation strategy)**

```
BEGIN
```

```
  Filter phishing emails from the dataset (emails labeled as "phishing")
  Apply TF-IDF to identify important terms and patterns unique to phishing emails
  Save the extracted phishing terms and their frequency to "phishing_signatures.csv"
  Update the phishing signature list as new phishing emails are identified
END
```

**6.Phishing email detection using signatures**

```
BEGIN
  Load phishing signatures from "phishing_signatures.csv"

  Function filter_email(new_email):
    Initialize counter for matching signatures to 0
    For each term in phishing_signatures:
      IF term in new_email THEN
        Increment counter
    END FOR

    IF counter > threshold (e.g., 3) THEN
      RETURN "Phishing Email Detected"
    ELSE
      RETURN "Legitimate Email"
    END IF
  END FUNCTION

  Check new incoming emails using filter_email function
END
```

**7.Continuous update and adaptation**

```
BEGIN
  Periodically update the phishing signatures list with new phishing emails
  Retrain SVM and XGBoost models with the updated dataset
  Continue refining the detection and mitigation process
END
```

## 4.3Mathmatical Explanation

### Support Vector Machine (SVM)

Support Vector Machine (SVM) is a supervised learning algorithm that is used for classification tasks. It works by identifying the optimal hyperplane that separates data points belonging to different classes. SVM is particularly effective in high-dimensional spaces and is known for its robustness against overfitting, especially when the number of features exceeds the number of samples.

### Mathematical Formulation:

D={(xi,yi)}i=i=$1^n$,where $x_i \in R^d$ is feature vector for the i-th sample, and $y_i \in -1,1$ is its corresponding label.The SVM seeks to find a hyperplane of the form:

$w^T x + b$=0

where w is the normal vector to the hyperplane, and b is the bias term.

To find this hyperplane, the SVM solves the following optimization problem:$min_b^w$1/2||w||$^2$

subject to the constraint:$y_i(w^T x_i + b)$ ≥1 $\forall_i$

This ensures that the margin between the classes is maximized, with the goal of achieving a clear separation.

$f_{svm}$(x)=$w^T$x+b ~>1

This function produces a score that represents the distance of the sample x from the separating hyperplane. The sign of the score indicates the predicted class, while the magnitude reflects the confidence of the prediction. In this model, these decision scores serve as input to the next stage of the pipeline.

**XGBoost (Extreme Gradient Boosting)**

XGBoost is an advanced implementation of gradient boosting for decision trees, designed to provide both efficiency and accuracy. It builds an ensemble of trees, where each new tree corrects the residuals (errors) of the previous trees, and is well-suited for tasks where high model performance is critical.

**Mathematical Formulation:**

For a given set of input features $x_i$,the XGBoost model predicts the output $\hat{y}_i$ by summing the contributions of K trees: $\hat{y}_i$=$\sum k = 1^k f_k(x_i)$ ~>2

where $f_k$ represents the k-th decision tree in the ensemble.

The training objective in XGBoost is to minimize a regularized loss function:

$\mathcal{L}(\emptyset) = \sum_{i=1}^{n} l(\hat{y}_i, y_i) + \sum k = 1^k \Omega(f_k)$ ~>3

Here,$l(\hat{y}_i, y_i)$ is the loss function that measures the difference between the predicted and actual values, and $\Omega(f_k)$ is a regularization term that penalizes complex models to prevent overfitting. The regularization term is defined as:

$\Omega(f_k) = \gamma T_k + 1/2\lambda\|w_k\|^2$ ~>4

where $T_k$ is the number of leaves in the k -th tree,$\gamma$ controls the complexity of the tree, and $\lambda$ is the regularization parameter for the tree weights.

Gradient Boosting Process:

XGBoost improves model predictions by iteratively fitting new trees to the residuals of previous trees. At each iteration t, the model updates its predictions as:

$\widehat{y_i^{(t)}} = \widehat{y_i^{(t-1)}} + \eta f_t(x_i)$ ~>5

where η is the learning rate, and $f_t(x_i)$ is the new tree added at iteration $t$ .

### 4.4Combined SVM + XGBoost Model

**Step 1: SVM as Feature Extractor**

The first step in the combined model is to train an SVM on the input dataset. For each input sample $(xi)$, the SVM produces a decision score $(f\text{SVM}(x_i))$ representing how confidently the sample belongs to a certain class. These decision scores are then used as features for the next model, XGBoost.

Mathematically, the new feature set is:

$$X\text{SVM} = \{f\text{SVM}(x1), f\text{SVM}(x2), \ldots, f\text{SVM}(x_n)\}$$

where each $(f_{\text{SVM}}(x_i))$ is the decision score for sample $(x_i)$.

**Step 2: XGBoost on SVM Scores**

The second step involves using XGBoost to classify the samples based on the decision scores from the SVM. These decision scores serve as the input features for the XGBoost model, which builds an ensemble of trees to improve prediction accuracy.

The prediction for each sample in XGBoost is given by:

$\hat{y}i = \sum k = 1^K f_k\big(f_{\text{SVM}}(x_i)\big)$ ~>6

where $(f_k)$ represents the $k$ -th decision tree, and the input feature is the SVM decision score.

**Final Prediction:**

The combined model's final prediction is based on the output of XGBoost, which leverages the SVM decision scores as input. This approach combines the linear classification power of SVM with the non-linear, ensemble-based learning of XGBoost, leading to improved classification performance.

## 4.5EVALUATION

To evaluate the performance of the combined model, we used several standard metrics: accuracy, precision, recall, F1-Score, and ROC-AUC. These metrics provide insight into how well the model can distinguish between phishing and legitimate emails.

[1] Accuracy: Accuracy is the proportion of correct predictions out of the total number of predictions made. It is calculated as:

Accuracy = (TP + TN) / (TP + TN + FP + FN)

Accuracy = 0.97

[2]Precision: Precision is the proportion of true positive predictions out of all positive predictions made by the model. It is calculated as: Precision = TP / (TP + FP)

Precision = 0.95

[3] Recall: Recall is the proportion of true positive predictions out of all actual positive cases in the dataset. It is calculated as:

Recall = TP / (TP + FN)

Recall = 0.98

[4] F1-Score: The F1-Score is the harmonic mean of precision and recall. It provides a balanced measure when both precision and recall are important.

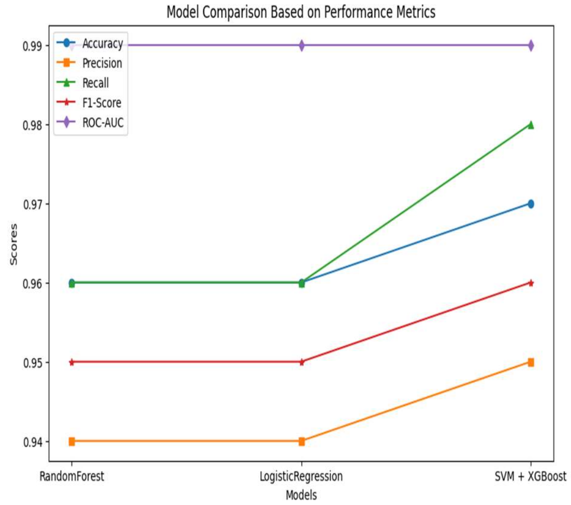
F1-Score = 0.96

[5] ROC-AUC: The ROC-AUC score is a performance metric that assesses the ability of the model to distinguish between classes. A higher ROC-AUC score indicates better performance.

ROC-AUC = 0.99

# 5.FINAL RESULT

This is a line chart representing the performance of three machine learning models: SVM + XGBoost, Random Forest, and Logistic Regression. These are shown on the X-axis while the Y-axis has been used to plot their performance scores. Lines, on the other hand, represent the different metrics like Accuracy, Precision, Recall, F1-score, and ROC-AUC.



From the chart, we can find:
SVM + XGBoost outperforms the other models on all metrics Random Forest and Logistic Regression perform similarly on most metrics with minor variations. Accuracy is usually high in general for all of the models, evidencing a good overall performance. Precision is also high, showing the models correctly identify the positive instances. While the recalls for Random Forest and Logistic Regression are somewhat lower than those for the SVM + XGBoost model, they probably miss some positive instances. F1-score is a balanced metric taking into account both precision and recall. SVM + XGBoost possesses the best F1-score, so it's the best in terms of balancing precision and recall. ROC-AUC is very high for all models, which means they are able to effectively discriminate between positive and negative instances. Overall, this chart shows that SVM + XGBoost is the best performing model for all metrics; however, this choice may depend on what the application requires. For example, in cases where high recall will be more important than high precision, it would be better to use Random Forest or Logistic Regression.

# 6.CONCLUSION AND FUTURE EXPANSION

## 6.1CONCLUSION

In this paper, we proposed the use of an integrated model of SVM and XGBoost for phishing email detection. Our integrated model effectively leveraged the power of both algorithms and showed a significant improvement in the detection of phishing emails. Other future directions may involve scaling this approach to real-time email filtering systems and exploring further methods of ensembling. Instead, with the extraction of phishing signatures, proactive measures can be taken against phishing attacks by finding common patterns within them and using those patterns as filters to bring education and training to the detection systems. This strategy is well within innovative, pragmatic ways in which a data-driven campaign may be leveraged in an ever-evolving cyber war against phishing.

## 6.2Future Expansion

The main extensions of this project in the future will involve key areas: improvement of phishing detection and mitigation strategies. First, there is the possibility of investigating the application of Natural Language Processing like BERT to improve the understanding of phishing emails. In tune with the process followed to feed them, advanced NLP models try to amplify the system's ability to detect minute patterns of language and ever-evolving phishing tactics that make phishing email classification robust.

Other futuristic developments could be real-time monitoring of mails and adaptive learning. This would be achieved by the continuous updating of the phishing signature database for newly identified phishing threats.

This will make the system proactive in detecting and blocking phishing attempts in real time, once integrated with the email infrastructure of an organization. Also, the adaptive learning model will improve with time-in preparation of feedback loops to retrain the model with new data, keeping it ahead of emerging threats.

Thirdly, this solution can be scaled up by detecting phishing threats across other channels of communication than just email; SMS, social media, and messaging apps among others are included herein. Thus, a multichannel phishing detection system would thereby place organisations in an advantageous position of protecting their users against an increasing array of different types of attacks.

Finally, the integration with cybersecurity awareness training platforms would add an educational element to the model. When the model has identified phishing emails in a user's inbox, real-time feedback would be given that should help the users' understanding of phishing risks and help develop a more cyber-aware workforce. This marrying of real-time detection, multi-channel threat identification, and user education would create an all-round adaptive, and forward-looking phishing mitigation solution.

# 7.REFERENCE

[1] Champa, A. I., Rabbi, M. F., & Zibran, M. F. (2024). *Curated Datasets and Feature Analysis for Phishing Email Detection with Machine Learning*. https://doi.org/10.1109/icmi60790.2024.10585821

[2] Priya, S., Gutema, D., & Singh, S. (2024). *A Comprehensive Survey of Recent Phishing Attacks Detection Techniques*. https://doi.org/10.1109/icitiit61487.2024.10580446

[3] Rajoju, R., Sathvika, V., Smaran, G. N. S., Tejashwini, C., & Reddy, G. A. (2024). *Text Phishing Detection System using Random Forest Algorithm*. https://doi.org/10.1109/icaaic60222.2024.10575110

[4] Champa, A. I., Rabbi, F., & Zibran, M. F. (2024a). *Why Phishing Emails Escape Detection: A Closer Look at the Failure Points*. https://doi.org/10.1109/isdfs60797.2024.10527344

[5] Gunjan, N., & Prasad, R. (2024). *Phishing Email Detection Using Machine Learning: A Critical Review*. https://doi.org/10.1109/ic2pct60090.2024.10486341

[6] Pullagura, L., Rao, D. M., Kumari, N. V., Lanke, R. K., Katta, S. K. G., & Chiwariro, R. (2024). *A Study of Suspicious E-Mail Detection Techniques*. https://doi.org/10.1109/idciot59759.2024.10467633

[7] Jain, N., Jaiswal, P., Sharma, S., Sharma, K., & Sharma, V. (2023). *A Machine Learning based Approach to Detect Phishing Attack*. https://doi.org/10.1109/icac3n60023.2023.10541835

[8] Jindal, N., Rastogi, D., Joshi, K., & Gupta, D. (2023). *Identification of Phishing Attacks using Machine Learning*. https://doi.org/10.1109/iciip61524.2023.10537706

[9] A, L. S. S., S, Y., & Jayapandian, N. (2023). *Machine Learning Based Spam E-Mail Detection Using Logistic Regression Algorithm*. https://doi.org/10.1109/ictbig59752.2023.10455970

[10] Divakarla, U., & Chandrasekaran, K. (2023). *Predicting Phishing Emails and Websites to Fight Cybersecurity Threats Using Machine Learning Algorithms*. https://doi.org/10.1109/smartgencon60755.2023.10442775

[11] Al-Yozbaky, R. S., & Alanezi, M. (2023). *A Review of Different Content-Based Phishing Email Detection Methods*. https://doi.org/10.1109/iec57380.2023.10438812

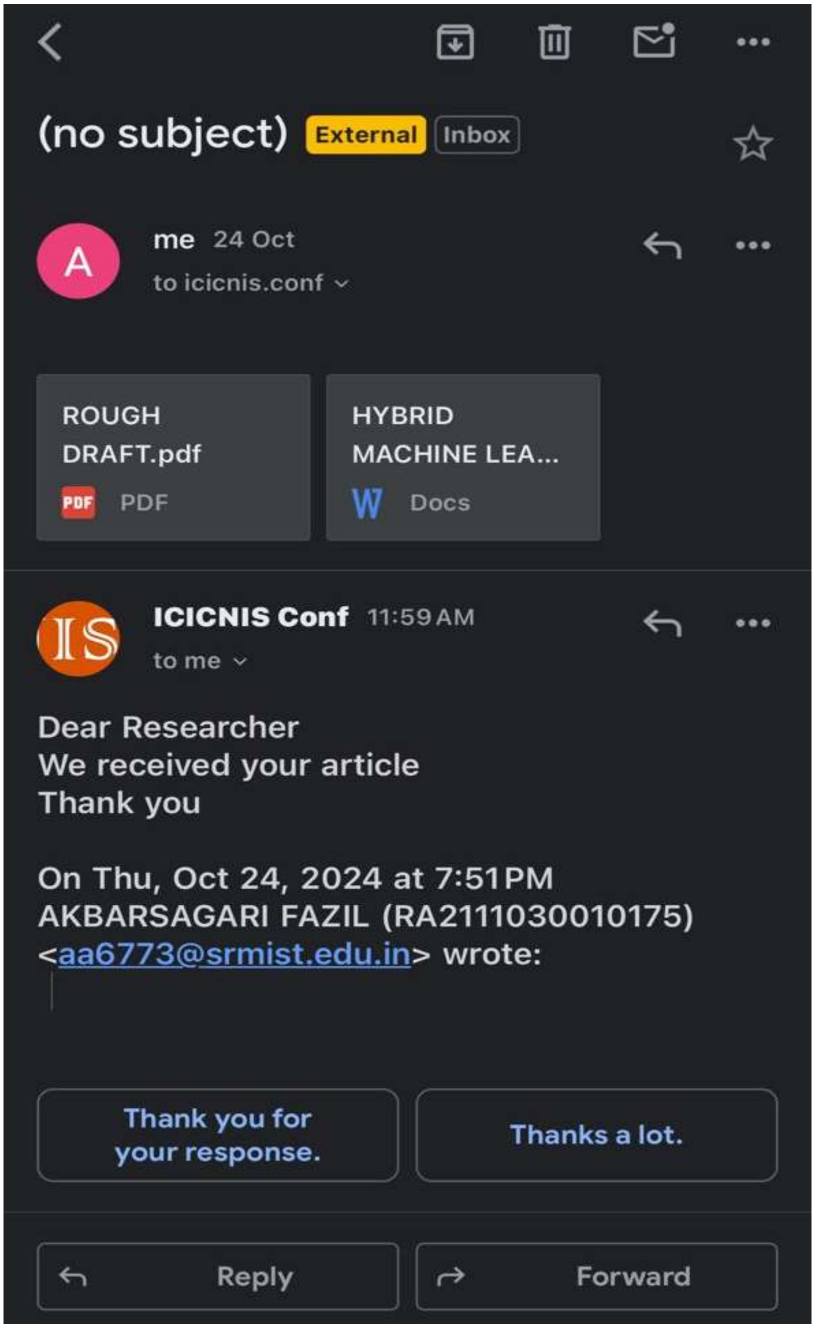
[12] *An Investigation into the Performances of the State-of-the-art Machine Learning Approaches for Various Cyber-attack Detection: A Survey*. (2024, May 30). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/10609847

[13] Priya, K. S., Chandrika, J. B., & Lakshmi, M. P. (2024). *Machine Learning-Based Phishing Website Detection A Comprehensive Approach for Cyber security*. https://doi.org/10.1109/icrtcst61793.2024.10578472

[14] Pullagura, L., Rao, D. M., Kumari, N. V., Lanke, R. K., Katta, S. K. G., & Chiwariro, R. (2024b). *A Study of Suspicious E-Mail Detection Techniques*. https://doi.org/10.1109/idciot59759.2024.10467633

[15] A. Chien and P. Khethavath, "Email Feature Classification and Analysis of Phishing Email Detection Using Machine Learning Techniques," *2023* https://doi.org/10.1109/CSDE59766.2023.10487729

# PAPER PUBLICATION STATUS



(no subject)

me 24 Oct
to icicnis.conf

ROUGH DRAFT.pdf

HYBRID MACHINE LEA...

ICICNIS Conf 11:59 AM
to me

Dear Researcher
We received your article
Thank you

On Thu, Oct 24, 2024 at 7:51PM AKBARSAGARI FAZIL (RA2111030010175) <aa6773@srmist.edu.in> wrote: