HYBRID MACHINE LEARNING MODEL TO DETECT AND MITIGATE EMAIL PHISHING ATTACKS

Akbarsagari
mohammad fazil
student,
Department of Networking and communications,
SRM Institute of Science & Technology, Kattankulathur,
Chennai aa6773@srmist.edu.in

Pallapolu Sai Vardhan
student
Department of Networking and communications,
SRM Institute of Science & Technology, Kattankulathur,
Chennai pz1371@srmist.edu.in

Jaini Eswar

student,
Department of Networking and communications,
SRM Institute of Science & Technology, Kattankulathur,
Chennai jj8322@srmist.edu.in

S Prabakaran,
Associate professor,
Department of Networking and Communications,
SRM Institute of Science & Technology, Kattankulathur,
Chennai
prabakes@srmist.edu.in

***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.**

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

I.INTRODUCTION

1.Hybrid Machine Learning Models SVM and XGBoost to Detect and Mitigate Email Phishing Attacks

Among other significant sources, the cybersecurity

threat includes attacks through e-mail phishing. Attackers try to steal sensitive information from victims by pretending to come from some legitimate organization. Due to such sophisticated phishing attempts, machine learning models have been in use for the detection and mitigation of phishing attacks. This paper proposes a hybrid approach by combining support vector machines with XGBoost for improving the detection accuracy in phishing emails and providing an additional layer of security. The combination of these two strong models significantly enhances the accuracy in classification by reducing the false positives, hence making this approach quite effective compared to the traditional methods available for phishing detection. This chapter discusses the concept of a hybrid machine learning approach and its realization and applications in the detection and mitigation of email phishing.

1.1 Introduction to Phishing Detection Using Machine Learning

Most phishing emails bypass traditional security measures because they're designed to appear as if they are real communications. Moreover, traditional phishing detection methods using blacklisting and heuristic-based filtering prove inefficient due to the increasingly sophisticated phishing methods. Machine learning models ensure a promising solution in learning the patterns and identifying the features that distinguish phishing emails from legitimate ones. The considered models, SVM and XGBoost, are widely used for classification problems. SVM is very efficient in binary classification problems, while XGBoost has strong performance using a gradient boosting algorithm. By integrating these two models, we can come out with a much more accurate phishing detection system[5].

1.2 Why Combine SVM and XGBoost for Phishing Detection

In this hybrid model, the strengths of both SVM and XGBoost are utilized to provide more accurate phishing detection. The general efficacy of SVMs in high-dimensional data, along with an obvious separation of classes, therefore provides a clear, optimized hyperplane which segregates the data points into distinct classes[8]. Conversely, XGBoost is based on ensemble learning methods since it improves the classification through multiple weak learners to come up with a strong model in classification. In order to make a more refined prediction and handle complex phishing attacks that may be evaded by a single model, the system followed the decision function of the SVM combined with the boosting technique of XGBoost[8].

1.3 Implementation and Evaluation

Certain crucial steps are required for the implementation of this hybrid model using SVM and XGBoost: The datasets of phishing and legitimate emails must be collected in a rich amount. Both classes should represent a somewhat equal distribution[2]. This dataset would need to undergo a rigorous preprocessing phase that involves cleaning, normalization, and feature extraction using TF-IDF.

Each of the models will then be trained on part of the dataset, which will save the rest for testing and validation. Performance metrics to be considered to evaluate the hybrid model will be accuracy, precision, recall, and F1-score, which give insight into how well the model can prevent phishing attempts while keeping the false positives as low as possible[3].

Second, this hybrid model is continuously updated with fresh data, considering that phishing techniques change a lot. As soon as newer phishing techniques begin to rise and surface, training the model again with fresh data lets it maintain an effectiveness rate that is high over time. This is very important in rendering cybersecurity effective amidst an ever-evolving landscape
Of threats.

This project will develop a robust, adaptive email phishing detection system to improve cybersecurity by combining the powers of SVM and gradient boosting using XGBoost. The solution addresses one of the more constant emerging dangers of the digital landscape and equips an organization with ways to keep sensitive information safe while reducing most risks that come with phishing attacks altogether.

The SVM is our first classifier, which learns the decision boundary on distinguishing phishing emails. In addition, XGBoost further refines predictions by incorporating SVM output feedback and achieving better detection results. Together, these components give a strong layered solution that can identify phishing emails while ensuring that the system remains both precise whilst still being adaptable to new and emerging phishing techniques[8].

II. LITERATURE SURVEY

A. I. Champa, M. F. Rabbi and M. F. Zibran, "Curated Datasets and Feature Analysis for Phishing 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[1].

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[15].

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.

A. 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[4]

.

Gunjan 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[5].

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.

S. Priya Dept. 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].

Nikhil Jindal Department of Computer Science & Engineering and Information Technology. The objective of this paper is to encompass dataset collection, pre-processing, machine learning model selection, and web-based application development for phishing URL detection. The overarching motivation lies in safeguarding users against the pervasive threat of phishing attacks. These attacks, prevalent and menacing, jeopardize personal and corporate data by exploiting private information. Existing defense measures, predominantly user awareness and education, often prove insufficient, as even informed individuals can succumb to sophisticated phishing tactics. Moreover, these attacks are increasing at an unprecedented rate and therefore, there is a need to deploy big data analytics to analyze the vast amount of information. Thus, the paper seeks to contribute to the field by providing advanced tools and methodologies capable of effectively identifying and thwarting these deceptive cyber assaults, offering enhanced protection to individuals and businesses in the digital realm[8].

"A Review of Different Content-Based Phishing Email Detection Methods" provides a thorough analysis of various

techniques used to detect phishing emails, highlighting the significant advancements in Artificial Intelligence (AI) and Machine Learning (ML). One notable development is the integration of Natural Language Processing (NLP) with ML, which has proven highly effective in phishing detection. Various features, such as contextual, syntactic, and semantic features, have been explored in past studies, with researchers like Vazhayil et al. utilizing classification techniques such as logistic regression, decision trees, random forests, and Support Vector Machines (SVM). Additionally, hybrid techniques, like those proposed by Hamid and Abawajy, combine content and behavioral features to improve detection accuracy. Bergholz et al. further categorized features into basic and advanced types, emphasizing the importance of both directly obtainable and latent features. However, the paper also highlights the limitations of NLP-based approaches, particularly their susceptibility to phishing emails that employ synonyms or varied phrasing, which can reduce detection effectiveness. A variety of detection methods are discussed, ranging from supervised techniques such as SVM, Logistic Regression, and Decision Trees to unsupervised approaches like k-means clustering and deep learning. The paper also references performance metrics from various studies, showcasing the high accuracy rates of methods like RRFST, C4.5, and CART. Additionally, the challenges of detecting phishing emails in non-English languages, especially Arabic, are explored, with the paper noting how linguistic complexity and rich vocabulary can impact NLP performance. Overall, the survey underscores the dynamic and evolving nature of phishing detection methods, emphasizing the need for continuous advancements to combat increasingly sophisticated phishing attacks.

lll. 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[2][13].

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[14][12].

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[11].

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[8].

IV.PROPOSED METHOD

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[5].

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[3].

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.

Activity diagram:

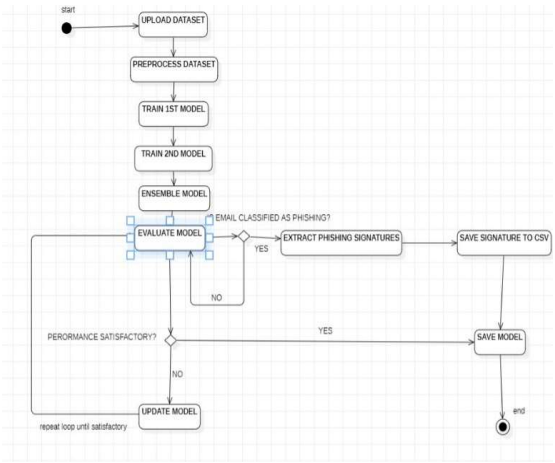

Figure.1

Figure 1: 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.

V.METHODOLOGY

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.

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.

A. Comparison 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.

B. PSEUDO 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
```

VI.MATHMATICAL 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:$y_i(w^T x_i + b) \geq 1 \ \forall_i$

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

f_{svm}(x)=w^Tx+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,γ controls the complexity of the tree, and λ 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 .

Combined 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:

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

VII.RESULTS AND DISCUSSION

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.

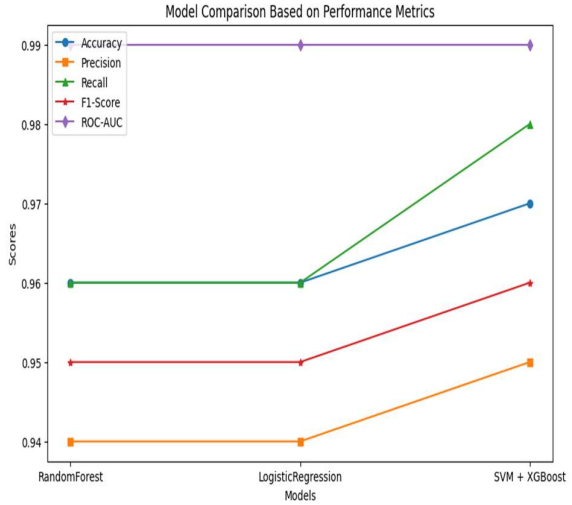

Figure.2

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.

VIII.FUTURE 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.

IX.CONCLUSION

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.

X.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