The Battle Against Phishing: Comparing Various Machine Learning Models for Email Threat Mitigation

Ahmet Emre Buz
Computer Engineering Department
Istanbul Technical University
Istanbul, Turkey
buz20@itu.edu.tr

Mustafa Can Çalışkan
Computer Engineering Department
Istanbul Technical University
Istanbul, Turkey
caliskanmu20@itu.edu.tr

Umut Ural
Computer Engineering Department
Istanbul Technical University
Istanbul, Turkey
uralu20@itu.edu.tr

Yusuf Emir Sezgin
Computer Engineering Department
Istanbul Technical University
Istanbul, Turkey
sezginy20@itu.edu.tr

***Abstract*—One of the most persistent and damaging threats in cybersecurity is phishing since attackers utilize ever sophisticated techniques to appeal to users and access critical data. Sometimes conventional detection methods such as spam filters and rule based systems are not enough against these developing challenges. Providing flexibility and improved detection abilities, machine learning (ML) models have become a potential answer to this issue. For the purpose of phishing email detection, this work evaluates numerous machine learning models—including Gaussian Naive Bayes, K-Nearest Neighbors, Logistic Regression, and Random Forest—thoroughly. Together with carefully produced datasets, we evaluate their performance using conventional assessment criteria including accuracy, precision, recall, F1-score. The results show that Random Forest and Logistic Regression preserve computational economy, have better recall and accuracy, and hence help to reduce false negatives. Approaching many aspects of phishing detection, Gaussian Naive Bayes and K-Nearest Neighbors show limitations. Our results highlight the need of model selection catered to application-specific objectives as well as the tradeoffs among accuracy, resource economy, and flexibility. We also identify as target points for next research important problems such adversarial resistance, cross-domain generalization, and real-time detecting. By means of advanced techniques and results from this study, researchers and practitioners can build strong and scalable defenses against the always growing phishing risk.**

1. Introduction

Phishing schemes are among the most often occurring and damaging forms of cybercrime in the digital era. Usually through well-crafted emails [1], people are duped into divulging confidential information including passwords and bank information. This hazard is far more severe given the massive volume of email contacts all around. There were estimated 347.3 billion emails sent daily as of 2023; 56.5% of these were labeled as spam [2]. Since they are becoming smarter [3], it is increasingly difficult to find scam emails using conventional techniques. Attackers constantly modify their strategies, so rule-based and content-dependent approaches become less dependeble and more difficult to implement in every context [4].

Since phishing attacks are inexpensive and significantly affect the people who engage in them, they are quite crucial. These assaults are simpler to execute than malware since they rely not on particular technical faults. They therefore complicate defenses" ability to maintain things safe [5]. Hacker activity accounted for more than 80% of hacks expected in 2022 90% of data breaches connected to these assaults [6]. Furthermore targeted more than others is the financial sector in the second quarter of 2023 alone, 23.5% of all phishing attempts occurred there [7].

Experts have turned to techniques based on machine learning (ML) and artificial intelligence (AI) to discover phishing emails more consistently [8]. While ML models may learn from data and adapt to fit attackers evolving plans [9], traditional approaches depend on set forms and patterns. Recent advances in machine learning (ML) and deep learning (DL) have made it feasible to design sophisticated systems for detection utilizing feature extraction, model selection and exhaustive training and testing strategies [10].

Though they offer a lot of promise, ML-based techniques have several flaws as well. Using the difficulty of creating models that are correct and applicable for

everyone, phishing emails are becoming more complex. Furthermore, the daily phishing attack targets of more than 100000 people worldwide highlight the need of developing efficient defenses [11]. Not enough traditional security tools—such as spam filters and rule-based heuristics—can halt phishing assaults, which are becoming more intelligent and frequent [12].

In this work, we fully examine many machine learning approaches for email phishing detection. We do several studies using actual data to gauge their performance. We especially examine in terms of accuracy, precision, memory, and generalizability how effectively these models handle the issues resulting from present phishing methods. Our contribution consists on three elements:

- **Systematic Evaluation:** We systematically evaluate multiple ML models to identify their strengths and weaknesses in detecting phishing emails under diverse conditions.
- **Performance Benchmarking:** We establish a performance benchmark for phishing detection, comparing traditional and ML-based methods using a robust dataset.
- **Practical Insights:** We provide actionable insights for researchers and practitioners to enhence the effectiveness of ML techniques in phishing prevention.

By looking at these important points, our study hopes to help the ongoing work to make defenses against the constantly changing threat of phishing attacks more reliable and scalable.

2. Related Work

Email phishing detection has been investigated by several researchers, and this field has benefited much from their efforts. Combining LSTM with CNN, Nijhum et al. developed HLSTMCNN, a hybrid deep learning model. The accuracy of phishing email detection improved notably with this model [13]. Through the benefits of both architectures, the proposed HLSTMCNN model achieves top performance. On Dataset 1 the model obtained an accuracy of 99.86% with a precision of 99.94%, a recall of 99.80%, and an F1 score of 99.84%. In the same manner, on Dataset 2 it attained an accuracy of 96.17%, higher than both CNN models and the single LSTM. Using LSTM, the hybrid architecture effectively gathers sequential information; CNN extracts local features; pre trained GloVe embeddings enhance semantic comprehension. By including more phishing emails to the sample, the model also addresses unbalanced datasets —a common occurrence in phishing detection. The findings reveal that HLSTMCNN offers a powerful and consistent approach for phishing detection in practical environments therefore defining a benchmark for further research in this field.

Rashed and Ozcan summarized in great depth machine and deep learning techniques for phishing detection. They underlined the benefits and drawbacks of many strategies [14]. According to their study, clever phishing detection methods are gaining favor. These comprise techniques like random forests, support vector machines (SVM), and neural network models including CNN and lSTM. They underlined the challenges resulting from shifting phishing techniques and the requirement of robust, flexible models able to handle attacks and shifting trends. The study revealed significant weaknesses in feature extraction techniques and some models lack resilience. This implies more sophisticated feature representation and hybrid architectures are necessary. This paper emphasizes the need of combining many machine learning models to solve the shortcomings of present techniques and increase the efficiency of phishing detection

In order to identify phishing, Asani and colleagues proposed a hybrid model combining the Gaussian Latent Variable Model (GLVM) [15] with the Support Vector Classifier (SVC). Their method produces remarkable results on a comprehensive dataset by combining the probabilistic feature extraction of GLVM with the high classification capacity of SVC. Emphasizing its prospective for efficient and light-weight phishing detection, the model displayed an accuracy of 98.43%, a precision of 99.28%, and a recall of 99.16%. To address significant problems such noise sensitivity and evolving phishing strategies, their method integrates machine learning with dimensionality reduction methods. This offers for cybersecurity uses a scalable and flexible solution.

Using Siamese deep learning, Remmide et al. demonstrated a novel approach to confirm authorship in phishing email detection [16]. Reaching an accuracy of 95% on the SeFACED dataset, their model groups emails based on stylistic and linguistic similarities. They demonstrated how one may leverage unique writing styles to distinguish phishing emails from actual ones by transforming phishing detection into an authorship verification issue. This approach demonstrates how well sophisticated neural network architectures may be used to enhance email security and lower phishing attempts.

Examining how conscientiousness and cue use interact in both controlled and real world settings [17], Williams and others investigated the psychological components of phishing detection. According to their findings, especially in controlled environments, increasing the accuracy of phishing detection may be much improved by applying additional clues. Furthermore, lower conscientiousness levels of participants often interacted with simpler phishing emails. This data emphasizes how prone someone is to fall for phishing schemes depending on their particular psychological features. It also emphasizes the requirement of certain treatments emphasizing on educating individuals to identify signs and raise their awareness of behaviors. Building on these initiatives, Atawneh and Aljehani developed a deep learning model especially for phishing

detection hence advancing the application of artificial intelligence in cybersecurity [18].

Several fresh trends and developments have driven this study. Alsuwaylimi's hybrid approach combining machine learning models with genetic algorithms demonstrates the significance of feature selection in certain languages, notably Arabic [19], for phishing detection.

Brindha and colleagues underlined the need of having intelligent deep learning models to provide scalable and robust phishing detection systems [20]. Proposed to identify and categorize phishing emails, the ICSOA-DLPEC model combines the Intelligent Cuckoo Search Optimization Algorithm with a Gated Recurrent Unit (GRU) model. Reaching a high accuracy of 99.72% the ICSOA-DLPEC model outperformed present techniques like CNN and LSTM models. The study underlined the need of effective feature extraction techniques like the N-gram approach and hyperparameter optimization to raise classification and detection performance. This methodology offers a decent structure for addressing the growing difficulties in cybersecurity related phishing campaigns. As shown by Bauskar et al, the application of big data analytics in artificial intelligence-driven phishing detection promotes the creation of scalable solutions for improved cybersecurity [21].

Deep learning models including BLSTM and FastText word embeddings have been proposed to identify phishing emails [22]. Applying this approach to an unbalanced dataset Wolert and Rawski obtained an accuracy of 99.12%, a precision of 98.43%, a recall of 99.49%, and an F1-score of 98.96%. The model underlined the need of having effective feature extraction for obtaining high detection rates by using preprocessing methods including tokenizing and stop-word elimination.

The structure HELPHED, which was created by Bountakas and Xenakis, is based on hybrid ensemble learning. Stacking and soft voting methods [23] are used in this system to combine content and textual features. HELPHED got an amazing F1-score of 0.9942 in a study that used 32051 good emails and 3460 fake emails. To deal with the complexity and changes in phishing emails, this work shows how important it is to use ensemble methods.

Stevanovič's work underlined the need of character and word embeddings in phishing detection [24]. Combining convolutional and recurrent layers on two public datasets let the model get an accuracy of 99.81% and an F1-score of 99.74%. This approach presents opportunities for more flexible detection systems as it illustrates how embedding-based approaches may learn significant characteristics on their own, independent of operator assistance.

Heiding et al. investigated how phishing detection and generation may benefit from large language models (LLMs) as GPT-4 [25]. Their studies revealed that, when combined with manual phishing systems like the V-Triad, LLMs obtained a click-through rate of up to 81%, far higher than conventional techniques. Furthermore, LLMs have shown promising results in spotting phishingulent emails occasionally, their accuracy surpasses human capacity. According to this study, LLMs have two crucial roles in phishing: they increase the sophistication of assaults and help to identify phishing efforts. It underlines how important balanced artificial intelligence development is to lower these hazards

Making phishing detection algorithms that are flexible, accurate, and computationally economical remains difficult even with recent advances. Good computational efficiency is given by lightweight heuristic approaches including those proposed by Jayaprakash et al. [26] and Bayesian spam filtering models [27]. They may fail, nevertheless to adequately address the evolving phishing strategies. Similarly, simpler techniques like rule-based incremental creation [29] and keyword matching [28] do not fit well to many phishing scenarios.

We examine these gaps by means of a systematic assessment system, considering the performance of many ML and DL models. With an eye on the tradeoffs between accuracy, interpretability, and efficiency, this paper provides doable suggestions for strengthening phishing email detection systems.

This work departs from other studies often stressing specific techniques, such as the ensemble classification model by Satheesh Kumar et al. [30] or Andriu's adaptive AI-based phishing detection approach [31]. Rather, it offers a detailed comparison of several ML and DL techniques. Our work is special as it evaluates the performance of deep learning models [18] under various conditions and incorporates the observations of Atawneh and Aljehani

The interaction of model correctness, computational efficiency, and adaptability is investigated in this paper It provides both scholars and practitioners a thorough and practical foundation.

3. Proposed Solution

The proposed solution implements many machine learning models and carefully evaluates their usefulness using a methodical and comprehensive approach to solve the phishing detecting problem. This work evaluates the performance of several well-known algorithms—Gaussian Naive Bayes, K-Nearest Neighbors (KNN), Random Forest, and Logistic Regression—in precisely identifying phishing emails while lowering important issues such false negatives, which pose significant cybersecurity risk. Emphasizing conventional measures like accuracy, precision, recall, and F1-score, the research forms a strict framework for

evaluating and contrasting model performance.

This work examines, for every model, the trade-offs between accuracy and recall. It is crucial as phishing detection systems typically struggle to reconcile high accuracy, which reduces false positives, and high recall, which guarantees that more phishing emails are properly caught. The study intends to provide specialized methods depending on the examination of these trade-offs, especially in cases when lowering false negatives comes first above enhancing accuracy.

The study also examines pragmatic elements that significantly influence the application of phishing detection mechanisms in actual life. This addresses processing speed, which evaluates system management of large datasets, and robustness to skewed datasets—a significant problem in phishing detection when legitimate emails exceed phoney ones. The research offers techniques to let the models run better, such as resampling the data and cost-sensitive learning, thus verifying how well they function in conditions where the chances are against them.

The comparison of the several machine learning models for phishing email detection offers useful understanding of their advantages and drawbacks. For some data distributions, for example, Gaussian Naive Bayes is computationally efficient and effective; yet, models such as Random Forest may offer better performance because of their capacity to manage complicated relationships among variables. When local data patterns are important, K-Nearest Neighbors might be helpful; but its computational expense in high-dimensional environments could restrict its use. Because of its simplicity and interpretability, logistic regression is a strong option for uses needing openness in decision-making.

This work advances our understanding of how to apply machine learning to improve email security by aggregating data to identify models that fit several applications. According to the study, we need a whole approach covering real-world application concerns as well as technological ones. This will help hacker detection systems reduce cyber hazards more successfully.

3.1. Methodology

Our suggested strategy to phishing detection is based on a clear, methodical methodology that guarantees consistently and dependably identification of bogus emails. Preprocessing carefully chosen email datasets [32] [33] [34] aiming to enhance the quality and usefulness of the data—was the first phase in the process. To guarantee that the dataset was correct and comprehensive, this stage comprised meticulously eliminating noise, mistakes, and extraneous objects. We also utilized feature extraction to identify and segregate the most significant elements of every email that is the subject line, sender information, and primary text.

We vectorized textual input to convert it into a format fit for machine learning models. These techniques made text-based elements numerical representations that preserved the semantic sense of the data therefore enabling their compatibility with machine learning methods. By closely analyzing the trends and correlations in the data, this phase guaranteed that the models could lay a strong basis for further research.

In awareness of the significant influence of dangerous URLs in bogus emails, we have included extra binary capability to identify the presence or absence of URLs inside the email text. This function was developed to add an additional layer of discriminative power to raise the accuracy of the model, thereby improving the vectorized data. Thus, by integrating this capability, our approach improved the robustness of the solution by tackling one of the most often used and dishonest tactics used by phishing adversary.

Four different machine learning models—Gaussian Naive Bayes (GNB), K-Nearest Neighbors (KNN), Random Forest, and Logistic Regression—were used in order to assess the efficacy of our approach. Every model was chosen for different advantages and features. The models were trained and tested on sets with training data ratios ranging from 20% to 90% of the whole dataset. This heterogeneity helped to highlight the scalability and adaptability of the models by means of an intensive performance analysis over several degrees of data availability.

We evaluated a thorough collection of important performance measures including accuracy, precision, recall, and the F1-score. These measures provide a complete assessment of every model's performance, assessing its capacity to minimize mistakes by precisely identifying phishing and authentic emails. We choose to investigate further the training data ratio that produced the best accuracy for every model.

We used a confusion matrix to investigate the performance of the best model more completely. The predictions of the model fell into four categories: true positives (phishing emails that were precisely identified), false positives (legitimate emails that were mistakenly marked as phishing) true negatives (legitimate emails that were correctly identified) and false negatives (phishing emails that were mistakenly classified as legitimate). This matrix gave the forecasts of the model a comprehensive viewpoint. This study highlighted the merits of the approach and offered important new perspectives on areas needing further improvement.

Finally, our methodical approach—which spans extensive preprocessing, feature engineering, a range of machine learning models, and a detailed performance analysis—helps

to highlight the dependability and efficacy of our phishing detection system. Through painstakingly addressing the complexity of phishing emails—including dangerous URLs and textual patterns—we provide a scalable and flexible platform fit for new risks. This finding marks a major turning point in the improvement of cybersecurity in an increasingly digital surroundings.

4. Analysis of Proposed Solution

4.1. Experiment Setup

The experimental setup for testing the proposed phishing detection solution was carefully designed to make sure that we could thoroughly and systematically assess how effective it is. The evaluation process was designed to mimic real world situations as much as possible to check how practical and applicable the solution is in general. To do this, we used publicly available email datasets from the real world, which show different features and qualities found in email communications.

The datasets went through a detailed preprocessing phase, which included two main steps: cleaning the data and extracting features. The data cleaning step included getting rid of noise, inconsistencies, and unnecessary information to make sure the data is of high quality. After that we did feature extraction to identify and measure important characteristics from each email. This helped the models use these characteristics for better learning and prediction.

The features that were extracted included vectorized representations of the subject line, sender details and the content of the emails. Also the presence or absence of a URL in the email body was recorded as a binary feature, because there is a strong link between malicious URLs and phishing attempts. The features gave a strong foundation for training and assessing the machine learning models.

We implemented four different machine learning models to evaluate the proposed solution: Gaussian Naïve Bayes (GNB), k-Nearest Neighbors (KNN), Random Forest, and Logistic Regression. The training data was divided into different proportions, from 20% to 90%, to study how the size of the training data affects the performance of each model. This detailed analysis helped to better understand how the availability of data affects the effectiveness of phishing detection systems

We evaluated the performance of each model using standard metrics like accuracy, precision, recall, and F1-score. These metrics gave a detailed view of the models' abilities, allowing for a strong comparison of their strengths and weaknesses.

To understand more about how the models work in practice, a confusion matrix was used for the model that performed the best, which was chosen based on its accuracy. The analysis of the confusion matrix showed the tradeoff between false positives and false negatives, providing useful insights into the operational challenges and risks that come with real-world deployment.

This detailed evaluation highlights the strengths and weaknesses of each model, offering useful insights for improving phishing detection systems. This study is important because it shows how to create stronger and more reliable email security systems.

4.2. Results

This part goes into great depth about how well different machine learning models work at finding fake emails. The main focus is on measures that come from the confusion matrices, with a focus on recall and false positives, since it's important to make sure that phishing emails are misclassified as little as possible. In this part, we also look at the pros and cons of each model and the trade-offs between their success measures.

4.2.1. Performance Metrics Analysis. Figures 1, 2, 3, and 4 show the performance metrics that is accuracy, precision, recall, and F1-score—for Gaussian Naive Bayes (GNB), K-Nearest Neighbors (KNN), Logistic Regression (LR), and Random Forest (RF), respectively, across varying training data proportions. The figures highlight the unique strengths of each model and their limitations in addressing the phishing detection problem.

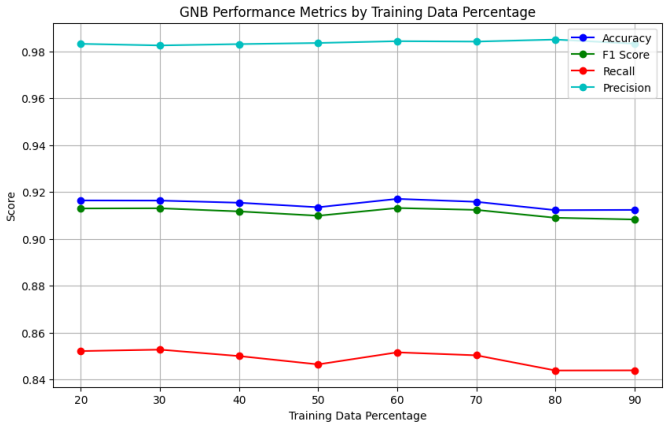

Figure 1: Performance metrics for Gaussian Naive Bayes across varying training data proportions.

Gaussian Naive Bayes, as shown in Figure 1, maintained stable performance across different training data proportions. However, its recall scores were slightly lower compared to Logistic Regresion and Random Forest. This indicates that it failed to identify some phishing emails, as evidenced by the higher number of false negatives in Figure 5. Despite its computational efficiency, GNB's reliance

on probabilistic assumptions about feature independence may have limited its ability to capture complex patterns in phishing emails.

As seen in Figure 2, the performance of KNN varied significantly based on the choice of k. Lower k values, such as $k = 3$, yielded higher recall scores but increased false positives due to the model's sensitivity to outliers. Larger k values improved precision at the expense of recall, highlighting the tradeoff between identifying phishing emails and avoiding false alarms. The results suggest that $k = 5$ provided the best balance, with competitive recall and reduced false negatives compared to lower k values.

Logistic Regression, showed in Figure 3, achieved consistently high recall and low false negatives across all training data proportions. This model's ability to accurately classify phishing emails coupled with its simplicity and efficiency, makes it a strong candidate for real-world deployment. Notably, Logistic Regression maintained its high recall even with smaller training data proportions, indicating robust generalization capabilities.

Random Forest, as shown in Figure 4, consistently delivered the highest recall and lowest false negatives among all models. This indicates its ability to capture complex patterns in phishing emails by leveraging ensemble learning. However the increased computational complexity of Random Forest may pose challenges in resource-constrained environments, making it less suitable for certain applications despite its superior performance.

4.2.2. Confusion Matrix Analysis. The confusion matrices and performance metrics for all models, evaluated at varying training data proportions, are shown in Tables 1, and 2. These matrices provide detailed insights into each model's performance in terms of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). Given the critical importance of minimizing false negatives for phishing detection, the matrices allow a direct comparison of how well each model performed under real-world conditions.

From confusion matrices of each model, key observations can be shown:

Gaussian Naive Bayes (GNB): As seen in Figure 5, GNB produced 67 false positives and 660 false negatives when trained with 90% of the dataset. While the number of false positives was relatively low, the high false negative count indicates that GNB frequently misclassified phishing emails as legitimate. This is a significant limitation in the context of phishing detection, where false negatives pose a critical risk.

KNN (k=5): Figure 6 shows the confusion matrix for KNN with $k = 5$, trained with 90% of the data. This model achieved a very low false negative count of 39, indicating

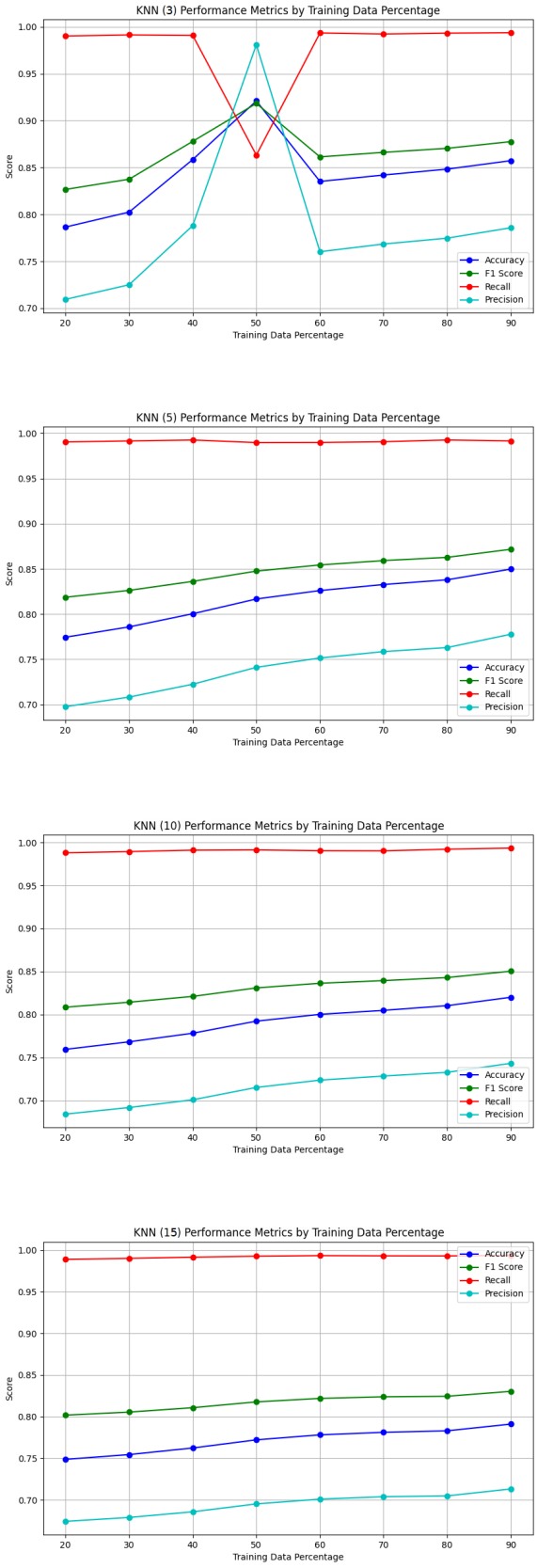

Figure 2: Performance metrics for KNN with varying k values.

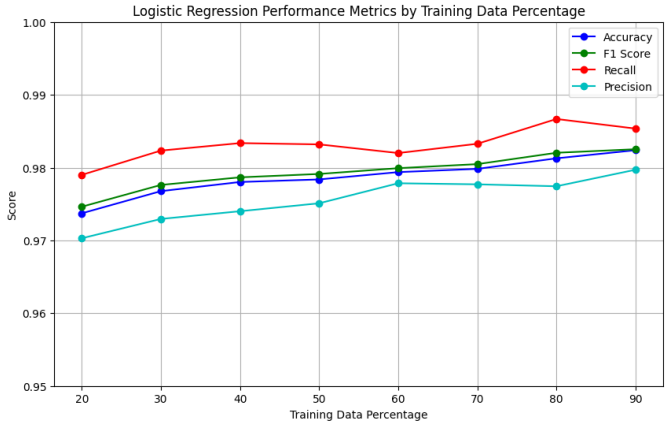

Figure 3: Performance metrics for Logistic Regression across varying training data proportions.

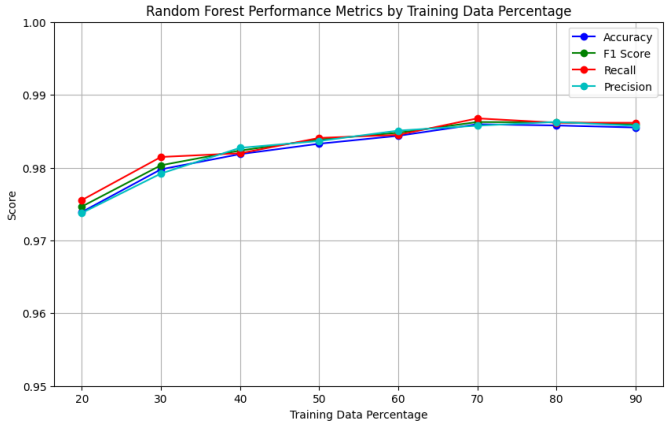

Figure 4: Performance metrics for Random Forest across varying training data proportions.

strong recall performance (0.991). However this came at the cost of 1,309 false positives, resulting in reduced precision. Such behavior highlights the sensitivity of KNN to outliers and its tendency to classify borderline cases as phishing, which may not be ideal for environments where precision is equally critical.

Logistic Regression (LR): The confusion matrix for Logistic Regression, shown in Figure 7, is based on 80% training data. This model delivered an excellent balance between false negatives (61) and false positives (56), achieving both high recall (0.987) and high precision (0.988) Logistic Regression's simplicity and efficiency, combined with its robustness against imbalanced datasets, make it a strong candidate for phishing detection.

Random Forest (RF): Figure 8 presents the confusion matrix for Random Forest trained with 90% of the data. Similar to Logistic Regression, this model achieved 56 false positives and 61 false negatives, coresponding to a recall of 0.987 and precision of 0.988. Random Forest's ability to capture complex patterns through ensemble learning contributed to its superior performance, albeit at a higher computational cost.

4.2.3. Tradeoffs and Model Comparisons. The results reveal significant trade-offs between memory and accuracy among the models. With few false negatives and high recall and accuracy, both logistic regression and random forest performed well in maintaining these parameters in balance. Conversely KNN demanded careful tuning to operate at its best and was sensitive to the value of k. On computers, Gaussian Naive Bayes was quick; yet, it was less helpful for identifying bogus emails as it struggled to lower false positives.

Reducing false negatives is rather crucial in real life as, should they be considered for real, scam emails can create serious security problems. Strong and with few false positives, logistic regression and random forest are suitable choices for high-security environments. Though we should be mindful of its shortcomings, Gaussian Naive Bayes might be a suitable option when resources are limited.

4.2.4. Real-World Implications. The findings of this study are important for instruments meant to spot phishings in practical environments. Low rates of false negatives in both Random Forest and Logistic Regression make them appropriate for uses demanding great dependability. Choosing a model requires one to take into account the boundaries of mathematics and the concessions between memory and precision. Finally, these models ought to be used solely in line with the particular program environment. This makes efficiency and performance really decent.

5. Discussion

The results of this work provide us with significant knowledge on the accuracy with which machine learning models can identify phoney emails at different times and levels. Here we discuss the meaning of the outcomes, the advantages and drawbacks of selecting a model, and how to use the model in the actual world.

5.1. Performance Insights

The study showed that Logistic Regression and Random Forest often performed better then other models in terms of accuracy, recall, and precision. These models demonstrated improved ability to detect phishing emails and decreased false negatives, which is very important for cybersecurity applications. It is simple and easy to understand, so Logistic Regression is a good choice. With this feature, it works especially well in places where honesty is key like in business or the law. On the other hand, Random Forest did a good job of finding complex patterns. This makes it a good choice when there are enough processing resources

and the goal is to make recognition work better.

Even though Gaussian Naive Bayes showed good accuracy, its high false negative rate makes it less reliable for detecting phishing. This limitation comes from its strict independence assumptions, which are often violated in the case of phishing emails where features interact in complex ways. The performance of KNN is greatly affected by how we choose the hyperparameter k. Even though we achieved high recall for some setups, the high false positive rate shows that we really need to adjust the parameters carefully.

5.2. Trade-Offs in Model Selection

Selecting the appropriate model for phishing detection involves balancing several factors:

- **Recall vs. Precision:** Reducing false negatives and ensuring that phishing emails are not mistakenly recognized as authentic ones depend critically on excellent recall. But as KNN shows, often attaining this target leads in a loss of accuracy. Recall should be given top priority in high-security surroundings. Conversely, circumstances involving a lot of emails might have to prioritize accuracy in order to minimize false positives' effect.
- **Computational Efficiency:** For systems with limited resources, logistic regression offers a fair compromise between how fast it can calculate results and how accurately it can detect outcomes. Although Random Forest is quite successful, in systems with low processing capability it might not be able to identify events in real time.
- **Scalability:** Large datasets need for models like Logistic Regression and Random Forest; KNN's computational cost rises significantly as the data quantity rises, therefore limiting its scalability.

5.3. Real-World Applicability

The results show how ımportant it is to use phishing detection models that are tailored to the needs of the target setting. Models like Logistic Regression are useful for organizations that do not have a lot of technical know-how because they are simple to use and understand. On the other hand, companies that are good with technology and have advanced computer resources might choose Random Forest because it has better abilities to identify patterns.

Flexibility is an important factor. As phishing techniques improve its important for a model to be able to apply what it has learned to new and unfamiliar data. Both Logistic Regression and Random Forest showed good performance even with unbalanced datasets, indicating that they can perform well in changing threat situations.

5.4. Limitations and Future Directions

Although the outcomes are favorable, several restrictions call for additional research even if it is already under progress. Although the datasets employed in this study are somewhat vast, they might not entirely reflect the spectrum of hacking methods applied in actual life. Future study should concentrate on aggregating a greater spectrum of relevant and varied datasets, including interactions in several languages and cultures setting.

Additionally examined in non-online environments were the models. This study did not investigate the issues of real-time phishing detection including latency and how to relate them to present email systems. We have to investigate light yet powerful application techniques like peripheral computation and model optimization if we are to make this work in real life.

Finally, systems employing machine learning still face a great threat from hostile assaults. Future research should aim to create phishing detection algorithms that are more impervious to efforts at tricking them. This will guarantee their performance even in the presence of clever attackers around.

5.5. Ethical Considerations

Especially with relation to user privacy, phishing detection systems have to be set up following standards. If we want to protect private email content and guarantee that model training works well, then approaches like federated learning and privacy-preserving machine learning should be looked at.

All things considered, this study presents a comprehensive overview of machine learning models employed for phishing detection combined with its pros and negatives as well as practical usage. By concentrating on new issues and resolving the already found ones, future research can assist to develop more adaptable, efficient, and safe email security solutions.

6. Future Work

Although this paper offers a thorough investigation of machine learning models for phishing email detection, some chances for future research can improve the general efficiency and applicability of phishing detection systems.

1) Emphasizing creative models like transformers and huge language models like BERT and GPT, future research should investigate the integration of sophisticated natural language processing algorithms. These algorithms shine in spotting the contextual and semantic subtleties sometimes

found in phishing emails.

2) Phishing attempts span many fields, languages, and disciplines. Essential are developing models that can efficiently generalize over several datasets without appreciable performance trade-offs. Achieving this can depend much on methods including domain adaptation and cross-domain transfer learning.

3) Adaptive feature engineering techniques are much needed as phishing techniques change quickly. Activity pattern tracking, metadata analysis, and real-time email traffic monitoring taken together would greatly improve detection capacity.

4) One looks to be a clear road to maximize strengths by combining numerous approaches. By hybrid systems integrating machine learning and deep learning or by ensemble classifiers matched with rule-based heuristics, one can enhance detection accuracy and recall.

5) Real-time detection systems need to be put into action and tested in real-world settings in order to see how useful the suggested methods are. This includes problems like delay, scalability, and working with email systems that are already in place.

6) Phishing detection systems have to be also strong against adversarial attacks. Models have to change to accommodate content modifications meant to evade detection methods as attackers always change their strategies.

7) Improved interpretability drives both regulatory compliance and user confidence. Explainable artificial intelligence (XAI) methods provide clear justifications for why an email is judged to be phishing or legitimate, therefore helping to foster transparency.

8) Dealing with phishing efforts in many linguistic and cultural settings calls for specific techniques. Development of multilingual models and datasets helps systems to be more widely applicable and efficient in diverse environments.

9) Dealing with personal email data calls for first consideration of privacy issues. By allowing models to be trained on dispersed datasets without sacrificing user privacy, techniques such as federated learning help to guarantee ethical and legal compliance.

10) Phishing detection systems can be improved by using knowledge on user behavior like response timings and click patterns. Content-based approaches mixed with behavioral analysis produces more strong and all-encompassing detection systems.

11) At last, one must take into account the harmony between operational needs, detection precision, and computational expenses. Limited resources companies can gain from looking at highly performing, reasonably priced phishing detection techniques.

By tackling these issues, future research can open the path for more scalable, flexible, dependable phishing detection systems that stay successful against always changing threats.

7. Conclusion

Phishing still causes a lot of problems in cybersecurity since attackers are getting better at fooling people and dodging cameras under monitoring. To identify phoney emails, many machine learning models—including Gaussian Naive Bayes, K-Nearest Neighbors, Logistic Regression, and Random Forest—were examined. We identified the main advantages and disadvantages of every model by carefully examining their performance with various degrees of training data.

Logistic regression and random forest turned out to be the best for balancing memory and accuracy. When you have to be quite sure you're not falling for phishing attempts, they are rather beneficial. Phishing emails with more complex patterns than Logistic Regression—which was more accurate and utilized less computer power—could be found better by Random Forest. Gaussian Naive Bayes struggled with complex hacking instances even if it was quick on computers. On the other hand, K-Nearest Neighbors responded sensitively to the selected values.

The suitable models for a given use case depend mostly on elements like available computing capability, the degree of accuracy needed, and the acceptance capacity of false positives or negatives. We also emphasized the importance of always improving these models to manage variations in hacking risk across time.

Phishing detection will be much improved in the future by means of advanced natural language processing techniques, models operating across various disciplines and languages, and deployment of real-time detection systems providing privacy protection. By addressing these issues, scientists can create trustworthy and scalable solutions to protect businesses and individuals from the growing risk of hacking incidents.

References

[1] P. Zhao and S. Jin, "Fewshing: A Few-Shot Learning Approach to Phishing Email Detection," *2024 IEEE 4th International Conference on Software Engineering and Artificial Intelligence (SEAI)*, Xiamen, China, 2024, pp. 371-375, doi: 10.1109/SEAI62072.2024.10674290. (Mustafa Can Çalışkan)

[2] J. Doshi, K. Parmar, R. Sanghavi, and N. Shekokar, "A comprehensive dual-layer architecture for phishing and spam email detection," *Computers & Security*, vol. 133, p. 103378, 2023, doi: 10.1016/j.cose.2023.103378. (Ahmet Emre Buz)

[3] K. Thakur, M. L. Ali, M. A. Obaidat, and A. Kamruzzaman, "A Systematic Review on Deep-Learning-Based Phishing Email Detection," *Electronics*, vol. 12, p. 4545, 2023, doi: 10.3390/electronics12214545. (Ahmet Emre Buz)

[4] R. Valecha, P. Mandaokar, and H. R. Rao, "Phishing Email Detection Using Persuasion Cues," *IEEE Transactions on Dependable and Secure Computing*, vol. 19, no. 2, pp. 747-756, Mar.-Apr. 2022, doi: 10.1109/TDSC.2021.3118931. (Ahmet Emre Buz)

[5] Q. Li, M. Cheng, J. Wang, and B. Sun, "LSTM Based Phishing Detection for Big Email Data," *IEEE Transactions on Big Data*, vol. 8, no. 1, pp. 278-288, Feb. 2022, doi: 10.1109/TBDATA.2020.2978915. (Ahmet Emre Buz)

[6] B. Gogoi and T. Ahmed, "Phishing and phishingulent Email Detection through Transfer Learning using pretrained transformer models," *2022 IEEE 19th India Council International Conference (INDICON)*, Kochi, India, 2022, pp. 1-6, doi: 10.1109/INDICON56171.2022.10040097. (Ahmet Emre Buz)

[7] N. Altwaijry, I. Al-Turaiki, R. Alotaibi, and F. Alakeel, "Advancing Phishing Email Detection: A Comparative Study of Deep Learning Models," *Sensors*, vol. 24, no. 7, p. 2077, 2024, doi: 10.3390/s24072077. (Ahmet Emre Buz)

[8] C. Thapa, J. W. Tang, A. Abuadbba, Y. Gao, S. Camtepe, S. Nepal, M. Almashor, and Y. Zheng, "Evaluation of Federated Learning in Phishing Email Detection," *Sensors*, vol. 23, no. 9, p. 4346, 2023, doi: 10.3390/s23094346. (Ahmet Emre Buz)

[9] N. Palanichamy and Y. S. Murti, "Improving Phishing Email Detection Using the Hybrid Machine Learning Approach," *Journal of Telecommunications and the Digital Economy*, vol. 11, no. 3, pp. 120–142, 2023, doi: 10.18080/jtde.v11n3.778. (Ahmet Emre Buz)

[10] Q. Qi, Z. Wang, Y. Xu, Y. Fang, and C. Wang, "Enhancing Phishing Email Detection through Ensemble Learning and Undersampling," *Applied Sciences*, vol. 13, p. 8756, 2023, doi: 10.3390/app13158756. (Mustafa Can Çalışkan)

[11] M. Somesha and A. R. Pais, "DeepEPhishNet: A deep learning framework for email phishing detection using word embedding algorithms," *Sadhana*, vol. 49, no. 3, Jul. 2024, doi: 10.1007/s12046-024-02538-4. (Mustafa Can Çalışkan)

[12] R. Lobo, M. N. Abbas, and M. N. Asghar, "Email Phishing Attack Detection using Recurrent and Feed-forward Neural Networks," *2023 Cyber Research Conference - Ireland (Cyber-RCI)*, Letterkenny, Ireland, 2023, pp. 1-6, doi: 10.1109/Cyber-RCI59474.2023.10671515. (Mustafa Can Çalışkan)

[13] N. A. Nijhum, Q. Li, and T. Yang, "HLSTMCNN: A Hybrid Deep Learning Model to Detect Phishing Email," *2023 3rd International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI)*, Wuhan, China, 2023, pp. 61-66, doi: 10.1109/CEI60616.2023.10528024. (Mustafa Can Çalışkan)

[14] S. Rashed and C. Ozcan, "A Comprehensive Review of Machine and Deep Learning Approaches for Cyber Security Phishing Email Detection," *International Journal of Scientific and Engineering Research (IJSER)*, vol. 3, no. 3, pp. 1–12, Sep. 2024. (Mustafa Can Çalışkan)

[15] E. O. Asani, V. O. Adedayo-Ajayi, A. E. Tunbosun, N. Enumah, and D. R. Aremu, "Detection of Phishing Emails using Support Vector Classifier and Gaussian Latent Variable Model," *2024 International Conference on Science, Engineering and Business for Driving Sustainable Development Goals (SEB4SDG)*, 2024, pp. 1–5, doi: 10.1109/seb4sdg60871.2024.10629863. (Mustafa Can Çalışkan)

[16] M. A. Remmide, F. Boumahdi, I. R. Ammar Aouchiche, and others, "A robust approach to authorship verification using siamese deep learning: application in phishing email detection," *International Journal of Speech Technology*, vol. 27, pp. 405–412, 2024, doi: 10.1007/s10772-024-10110-y. (Mustafa Can Çalışkan)

[17] R. Williams, B. W. Morrison, M. W. Wiggins, and P. Bayl-Smith, "The role of conscientiousness and cue utilisation in the detection of phishing emails in controlled and naturalistic settings," *Behaviour & Information Technology*, vol. 43, no. 9, pp. 1842–1858, 2023, doi: 10.1080/0144929X.2023.2230307. (Umut Ural)

[18] S. Atawneh and H. Aljehani, "Phishing Email Detection Model Using Deep Learning," *Electronics*, vol. 12, p. 4261, 2023, doi: 10.3390/electronics12204261. (Umut Ural)

[19] A. A. Alsuwaylimi, "Enhancing Arabic Phishing Email Detection: A Hybrid Machine Learning Based on Genetic Algorithm Feature Selection," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 15, no. 8, 2024, doi: 10.14569/IJACSA.2024.0150832. (Umut Ural)

[20] R. Brindha, S. Nandagopal, H. Azath, V. Sathana, G. P. Joshi, and others, "Intelligent deep learning based cybersecurity phishing email detection and classification," *Computers, Materials & Continua*, vol. 74, no. 3, pp. 5901-5914, 2023, doi: 10.32604/cmc.2023.030784. (Umut Ural)

[21] S. R. Bauskar, C. R. Madhavaram, E. P. Galla, J. R. Sunkara, and H. K. Gollangi, "AI-Driven Phishing Email Detection: Leveraging Big Data Analytics for Enhanced Cybersecurity," *Library Progress International*, vol. 44, no. 3, pp. 7211-7224, 2024, Available at SSRN: https://ssrn.com/abstract=4975526. (Umut Ural)

[22] R. Wolert and M. Rawski, "Email Phishing Detection with BLSTM and Word Embeddings," *International Journal of Electronics and Telecommunications*, pp. 485–491, 2023, doi: 10.24425/ijet.2023.146496. (Umut Ural)

[23] P. Bountakas and C. Xenakis, "HELPHED: Hybrid Ensemble Learning PHishing Email Detection," *Journal of Network and Computer Applications*, vol. 210, p. 103545, 2023, doi: 10.1016/j.jnca.2022.103545. (Umut Ural)

[24] N. Stevanović, "Character and Word Embeddings for Phishing Email Detection," *Computing and Informatics*, vol. 41, no. 5, pp. 1337–1357, 2022, doi: 10.31577/cai_2022_5_1337. (Umut Ural)

[25] F. Heiding, B. Schneier, A. Vishwanath, J. Bernstein, and P. S. Park, "Devising and Detecting Phishing Emails Using Large Language Models," *IEEE Access*, vol. 12, pp. 42131–42146, 2024, doi: 10.1109/access.2024.3375882. (Yusuf Emir Sezgin)

[26] J. R. Jayaprakash, K. Natarajan, J. A. Daniel, C. V. Chinnappan, J. Giri, H. Qin, and S. Mallik, "Heuristic machine learning approaches for identifying phishing threats across web and email platforms," *Frontiers in Artificial Intelligence*, vol. 7, p. 1414122, 2024, doi: 10.3389/frai.2024.1414122. (Yusuf Emir Sezgin)

[27] M. Sharabov, G. Tsochev, V. Gancheva, and A. Tasheva, "Filtering and Detection of Real-Time Spam Mail Based on a Bayesian Approach in University Networks," *Electronics*, vol. 13, p. 374, 2024, doi: 10.3390/electronics13020374. (Yusuf Emir Sezgin)

[28] K.-T. Tham, K.-W. Ng, and S.-C. Haw, "Phishing Message Detection Based on Keyword Matching," *Journal of Telecommunications and the Digital Economy*, vol. 11, no. 3, pp. 105–119, 2023, doi: 10.18080/jtde.v11n3.776. (Yusuf Emir Sezgin)

[29] M. SatheeshKumar, K. G. Srinivasagan, and G. UnniKrishnan, "A lightweight and proactive rule-based incremental construction approach to detect phishing scam," *Information Technology and Management*, vol. 23, pp. 271–298, 2022, doi: 10.1007/s10799-021-00351-7. (Yusuf Emir Sezgin)

[30] A. S, P. R. Nishant, S. Baitha, and K. D. Kumar, "An Ensemble Classification Model for Phishing Mail Detection," *Procedia Computer Science*, vol. 233, pp. 970–978, 2024, doi: 10.1016/j.procs.2024.03.286. (Yusuf Emir Sezgin)

[31] A.-V. Andriu, "Adaptive Phishing Detection: Harnessing the Power of Artificial Intelligence for Enhanced Email Security," *Romanian Cyber Security Journal*, vol. 5, no. 1, pp. 3–9, 2023, doi: 10.54851/v5i1y202301. (Yusuf Emir Sezgin)

[32] A. I. Champa, M. F. Rabbi, and M. F. Zibran, "Why phishing emails escape detection: A closer look at the failure points," in *12th International Symposium on Digital Forensics and Security (ISDFS)*, 2024, pp. 1–6. (Yusuf Emir Sezgin)

[33] A. I. Champa, M. F. Rabbi, and M. F. Zibran, "Curated datasets and feature analysis for phishing email detection with machine learning," in *3rd IEEE International Conference on Computing and Machine Intelligence (ICMI)*, 2024, pp. 1–7. (Joint)

[34] "Phishing Email Data by Type," Kaggle, Apr. 7, 2022. [Online]. Available: https://www.kaggle.com/datasets/charlottehall/phishing-email-data-by-type. (Joint)

8. Appendix

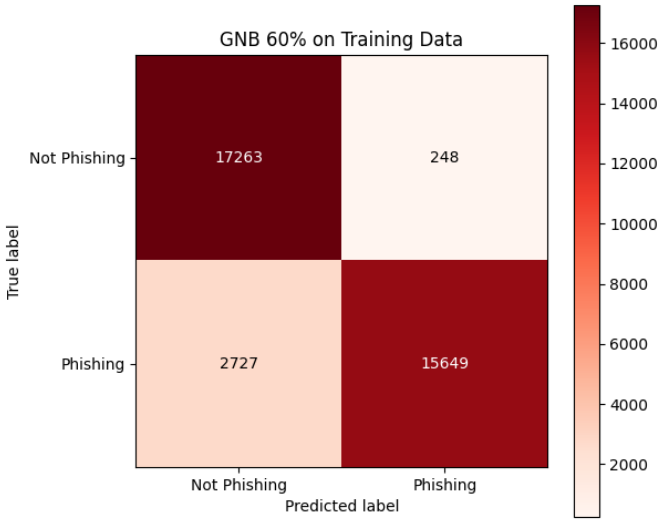

Figure 5: Confusion matrix for the Gaussian Naive Bayes (GNB) model.

TABLE 1: Confusion Matrix for All Models (90% Training Data)

Model	TP	TN	FP	FN
Gaussian Naive Bayes	15649	17263	248	2727
KNN (k=5)	4579	3045	1309	39
Logistic Regression	4442	4372	92	66
Random Forest	13568	12969	196	182

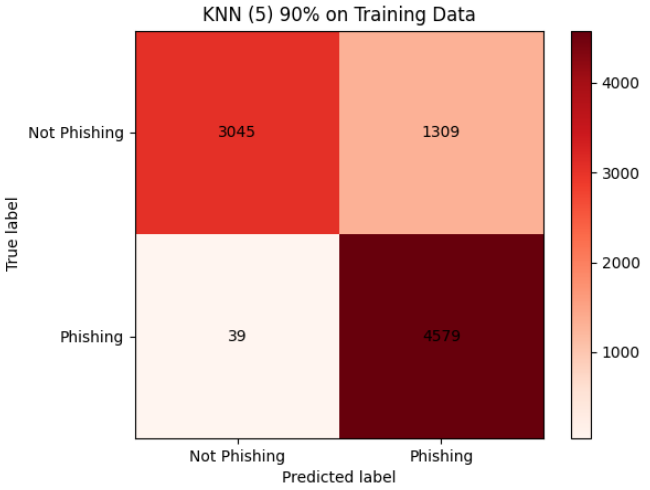

Figure 6: Confusion matrix for the K-Nearest Neighbors (KNN) model with k=5.

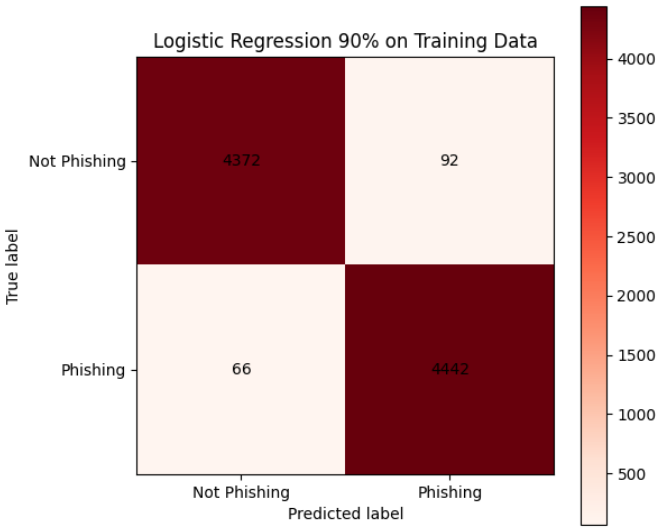

Figure 7: Confusion matrix for the Logistic Regression model.

TABLE 2: Performance Metrics for All Models (90% Training Data)

Model	Recall	Precision	Accuracy	F1-score
Gaussian Naive Bayes	0.852	0.984	0.917	0.913
KNN (k=5)	0.992	0.778	0.850	0.872
Logistic Regression	0.985	0.980	0.982	0.983
Random Forest	0.987	0.986	0.986	0.986

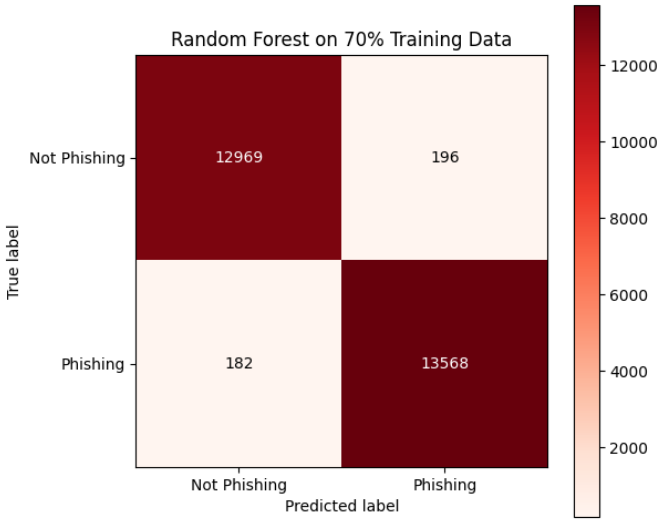

Figure 8: Confusion matrix for the Random Forest model.