International Journal of Computing Sciences Research (ISSN print: 2546-0552; ISSN online: 2546-115X)
Vol. 7, pp. 2358-2383
doi: 10.25147/ijcsr.2017.001.1.165
https://stepacademic.net

Long Paper

Email Attacks: An Ensemble Algorithm Utilizing Machine Learning for Phishing Detection Towards Potential Attack Prevention

Erwin E. Guerra
College of Computer Science, University of Makati, Philippines
ORCID: 0000-0003-3286-6661
erwin.guerra@umak.edu.ph
(corresponding author)

Date received: June 16, 2023
Date received in revised form: July 19, 2023; July 29, 2023
Date accepted: August 5, 2023

Recommended citation:

Guerra, E. E. (2023). Email attacks: An ensemble algorithm utilizing machine learning for phishing detection towards potential attack prevention. *International Journal of Computing Sciences Research*, 7, 2358-2383. https://doi.org/10.25147/ijcsr.2017.001.1.165

Abstract

Purpose – This study is designed to validate the effectiveness of the ensembled algorithm of two machine learning algorithms in the detection and potential prevention of email intrusion in corporate firms, government institutions, and individuals as compared to other studies that use only a single selected best machine learning for email detection and filtering.

Method – The sampling method utilized the best algorithms for the ensemble which are Random Forest and Support Vector Machine (SVM) and were trained on the Kaggle dataset. SVM was embedded in the designed web page for email spam detection, while Random Forest was implemented in a browser extension for the detection and prediction of phishing links in emails.

Results – The test results showed that both algorithms achieved high accuracy rates, with SVM achieving an accuracy of 0.97% and Random Forest achieving an accuracy of 0.87%. As an ensemble approach, Random Forest and SVM advance if not outclass them in terms of accuracy, precision, recall, f1 score, true positive rate, and false positive rate.

Conclusion – From the findings, this study suggests that ensembled machine learning algorithms can be effective in detecting spam and malicious links in emails. The high accuracy rates achieved by both models indicate that they can be used as reliable ensembled tools for email threat detection and security.

Recommendations – It is highly recommended to embed the model system or the like into several email providers to automatically detect spam without having to copy and paste the email content into a webpage. Also, disabling malicious links and detecting malicious email attachments (payloads) should be included to further the capabilities of this study.

Theoretical Implications – The study on ensembled algorithms in machine learning if carefully selected will surely advance the accuracy detection of false positives or false negatives in email. This will lead to trust and worry-free email usage for everyone.

Keywords – machine learning, ensemble algorithm, email attacks, phishing, malicious links detection, attack prevention

INTRODUCTION

Email fraud and scam messages are widespread nowadays not only in the Philippines but around the world. This is due to communication requirements that most professional communications and transactions are done through email. These online email services are the most convenient way of doing professional and non-professional communication as compared to the traditional way of doing so. This is the obvious reason why attackers turn to target emails by sending payloads and malicious links aside from compelling email content to persuade the users.

While email has been one of the most convenient methods of communication, utilizing emails carries a lot of advantages since most of its services are free to use, allowing users and firms to make use of this efficient way of communication for business purposes.

However, though existing email filtration and detection occur in email provider's server, the continuous growth of phishing attacks still cost users, and firms large amounts of losses because of the attackers recognized loophole penetrating the system. Government, Banks, and established institutions were victims of these attacks. This is due to some factors that users or employees of an organization were deficient in the awareness and practices on preventing and identifying malicious emails that can lead to such attacks. Furthermore, detecting phishing threats in email content is crucial to maintain the organizational operation running and fending against attacks such as data breaches, phishing, and financial loss. Moreover, utilizing phishing detection is essential to prevent all hackers or cybercriminals from infiltrating or attacking users and organizations. Integrating a more systematic flow of solutions to identify phishing attacks in email content could improve attack prevention.

The utilization of an ensemble machine learning algorithm in this paper as an introduced solution through the author's simply developed web application for email content and malicious link verification has the competence to advance the detection process against malicious messages in a particular email. Thus, testing the five best-performing classifiers for email spam filtering which are Random Forest, Support Vector Machine, K-Nearest Neighbors, Naïve Bayes, and Decision Tree could support in choosing the two top algorithms for the ensembled model. This ensemble study makes it easier for the email system to identify potential threats feasibly better than those existing studies selecting only one best algorithm to do the task.

LITERATURE REVIEW

Phishing Detection

Following the study of Alani et al. (2022), phishing is one of the most often used tactics for carrying out cyberattacks and is continually expanding. 97% of users, as shown by recent statistics, are unable to identify a sophisticated phishing email. Traditional approaches such as rule-based filters and legacy blacklists are no more effective to reduce the rising hazards and complexity of phishing's level since more than 1.5 million recently created fraudulent websites are generated each month. In this research, the researchers introduce PhishNot, a solution for identifying phishing URLs based on machine learning techniques. As a result, their work predominantly employs a "learning from data" driven methodology that is tested against a representative situation and dataset.

On the other hand, the study by Ouyang (2022), states that blacklist technologies are often used to prevent malware threats on emails, however, research revealed that they have limitations including inadequate coverage and a delayed reaction to new threats.

Concerning the research paper of Lee and Lee (2022), due to the COVID-19 pandemic, malware attacks are causing more harm than before. Examples include ransomware and spear phishing attacks on enterprises or institutions. Since malware attacks primarily use email as their major method of penetration, several research papers have been carried out to mitigate the occurrence of malicious emails in the work environment.

In agreement with Alkhalil et al. (2021), phishing is one instance of a very efficient type of digital crime that enables attackers to trick users and acquire confidential data. Phishing attacks have the potential of inflicting substantial damage on their victims, including the theft of private information, identities, businesses, and state secrets.

According to the paper of Wei et al. (2020), the evolution of the Internet has been accompanied by advancements in fraud techniques and strategies for obtaining sensitive information about individuals, including logins and passwords. In this article, the researchers demonstrate how convolutional neural networks can be used to identify fraudulent URL addresses with nearly 100% accuracy.

According to Broadhurst and Trivedi (2020), the Australian Communications and Media Authority's Spam Intelligence Database contained a subset of ten percent (10%) of spam email samples from a dataset of 25.76 million emails in 2016. These emails were scanned with the use of VirusTotal. The result revealed that one in ten percent (10%) of emails were found to have malware infiltrated, and nine percent (9%) of emails were found to be inactive. Approximately 31.8% of the compromised websites, which amounts to 81,176 sites, were identified as being involved in phishing, 58.4% were compromised by a type of malware known as a trojan, and 40.6% were websites specifically created for malicious purposes. 115, 025 attachments were also scanned which showed that 36,405 were infected with numerous malware types. The most prevalent malware was a variety of trojans and ransomware.

As explained by Sohail (2021) in their published research, states that antivirus software is unable to detect malware in a file because of its file format. Malware is often hidden within the macro of the Word document, and it is usually sent through emails. Rich text files are not executable files; hence, they can then bypass all the security of antivirus software. The researchers, therefore, proposed a system that automatically detects and analyses Microsoft Word files using Python.

As per Wu & Guo (2022) in their study, organization security is significantly affected by the email threat in today's era, which includes a variety of fraudulent situations including phishing, fraud, blackmail, and malvertising. Traditionally, to screen out malicious emails that rely on malicious words in the content of an email, enterprises require to maintain a greylist. Unfortunately, there are new techniques that hackers develop to bypass the traditional anti-spam.

Baig (2021) argued that one of the biggest problems with the Internet is email spam, which annoys consumers and harms enterprises. To prevent spam, one of the traditional methods is filtering. Email filters are frequently used to sort through incoming mail, safeguard computers against malware, and get rid of spam. For that reason, the researcher proposed this method for classifying unsolicited emails utilizing Support Vector Machines which outperformed other classifiers. The Support Vector Machines yielded an accuracy of 97.29%, while Decision Tree Classifier yielded 86.24%, Naïve Bayes yielded 85.60%, and BernoulliNB yielded 80.20%. Using the outcomes of this accuracy test, the best algorithm for the project was determined. This study contends that using SVM to identify spam emails produces more accurate spam detection results.

As discussed by Karim et al. (2019), the rapid proliferation of phishing emails, in the form of spam, spear phishing, or malware delivered via email, has led to a growing demand for dependable and advanced email filters that are capable of filtering spam emails. This research article outlines a concentrated literature review of AI (Artificial Intelligence) and ML (Machine Learning) approaches to detect spam emails intelligently, which the researcher argues can aid in the development of effective countermeasures.

As discussed by Zhang et al. (2019), Targeted Malicious Email (TME) is being used as a threat vector on the Internet nowadays. Targeted Malicious Emails attack often uses the personal information of an organization's employee or employer to make the email appear more convincing and believable. Such emails often contain malicious URLs or attachments that can cause severe damage to an organization. In addition, it focuses on compromising the security of an organization to access crucial information.

Zhang et al., (2020) proposed dynamic detection techniques to efficiently address this novel form of email-based cyber-attack. The researchers simulate the opening of a malicious URL and attachment in email using Virtual Machine. They also used Memory Forensics Analysis and Virtual Machine Introspection for them to get the real-time attributes of the content of an email. After that, the researchers used the AdaBoostM1 ensemble learning technique and a combination of Voting to detect Targeted Malicious Email attacks.

Machine Learning

According to Akhtar and Feng (2022), a new type of harmful software known as polymorphic malware is more versatile than viruses from previous generations. To prevent being recognized by traditional signature-based malware detection algorithms, polymorphic malware frequently alters its characteristic properties. Therefore, researchers utilized a diverse range of machine-learning approaches to detect such dangerous threats or viruses. Different techniques were utilized, including Random Forest, Naive Bayes, SVM, Decision Tree, CNN, and J48. The result showed that the performance of the Decision Tree yielded the highest detection accuracy of 99%, surpassing all other techniques.

Also, Ahmed et al. (2022) stated that email spam, otherwise referred to as unwanted or unsolicited email, is a sort of email that can be utilized to negatively impact users by wasting their time and attempting to obtain their personal information. Today's email and IoT service providers face significant and massive challenges with spam identification and filtration. Filtering email is one of the most important and well-known methods available for identifying and avoiding spam among all the methods now in use. This study classifies machine learning techniques applied to spam filtering methods and investigates their applications in both email and IoT platforms.

Likewise, Ojewumi et al. (2022) stated that the paper's rule-based approach to detecting phishing involved training three machine learning models on a dataset that contained fourteen (14) different features. The three distinct machine learning algorithms used were Random Forest, Support Vector Machine (SVM), and k-Nearest Neighbor (KNN) with Random Forest yielding the best result for detecting phishing.

Based on the study of Alhogail and Alsabih (2021), phishing emails are successful in deceiving people by exploiting their emotions and creating a sense of urgency, making them believe that immediate action is required, resulting in substantial monetary and data losses. As a result, humans need more efficient and automatic phishing detection techniques because they cannot simply rely on people to identify phishing. In this study, the researcher proposed a phishing email classifier model utilizing Graph Convolutional Network (GCN) and Natural Language Processing (NLP). The literature has demonstrated the success of GCN in categorizing text, and their work has demonstrated it to be successful in enhancing email phishing detection precision.

As reported by Haiba and Mazri (2021), IoT devices are the future, they will transform every aspect of our lives, beginning with smart homes, businesses, and e-healthcare systems. Furthermore, as IOT networks are being used more and more, threats are constantly updating their use of these technologies to exploit their flaws. Malware has multiplied and found several methods to break through, thus it is now more important than ever to have an effective malware detection system that can keep up with them as they grow stronger.

The study of Ding et al. (2021), proposes a machine learning-based approach as an alternative and efficient method to differentiate spear phishing emails. To achieve this, the study used 21 stylometric features extracted from emails, three forwarding features from an Email Forwarding Relationship Graph Database, and three reputation features obtained from two third-party threat intelligence platforms; VirusTotal and Phish Tank will be extracted. After that, an improved Synthetic Minority Oversampling technique algorithm was made to mitigate the effects of imbalanced data. Lastly, to recognize spear phishing emails from non-spear phishing emails, the researcher used four machine learning algorithms. The dataset of the researcher comprises 417 spear phishing emails and 13,916 non-spear phishing emails. The researcher attained a maximum recall rate of 95.56%, a precision rate of 98.85%, and an F1-score of 97.16% by utilizing forwarding features, reputation features, and Synthetic Minority Oversampling Technique.

Likewise, Khan et al. (2021) disclose that based on the Internet Security Threat Report for 2019 by Symantec, Microsoft Office files comprised 48% of all malicious email attachments in the year 2018. In this paper, the researcher provides a technique with high accuracy for identifying malicious office files. Using a Random Forest classifier with a static analysis and dynamic strategy which is called the hybrid method, the researcher was able to attain a detection accuracy of 99.57%, which was the highest among all the experiments conducted.

Siddique et al. (2021) argued in their research that email is widely utilized as a means of social communication, utilized in both formal and informal contexts. Unwanted and improper emails, or spam, are frequently sent to compromise security. These emails contain phishing URLs, promotions, and commercial content, and are typically sent to many recipients selected at random. The proposed research utilized Naïve Bayes, CNN, SVM, and LSTM among other available machine learning techniques to identify and categorize the content of an email. The findings revealed that the LSTM model exhibited superior performance compared to the other models, achieving a maximum accuracy score of 98.4%.

In conformity with Bawazeer et al. (2021) on their study, Machine Learning (ML) algorithms have been using Hardware Performance Counters (HPCs) events more and more over the past ten years to detect malware. This research introduces an analytical study to classify the HPC-based machine learning methods utilized for malware detection. Moreover, a variety of studies from the literature are simulated using the Neural Network (NN) approaches, such as the Multi-Layer Perceptron (MLP), Full Order Radial Basis Function (RBF), and Convolutional Neural Network (CNN) techniques. The simulation results indicate that MLP, Full Order RBF, and CNN have accuracy values of 96.95%, 98.22%, and 98.68%.

In the study of Quang et al. (2021), the emergence of the big data era can be attributed to the quick development of new technologies, such as smart gadgets, 5G connectivity, and other smart devices. Machine learning faces several difficulties as a result of big data, particularly in the area of phishing detection. This study intends to give a synthesis and evaluation of recent research on utilizing machine learning-based phishing detection for big data. 30 publications from various journals and conference proceedings were critically reviewed as part of this study's systematic literature review (SLR) methodology.

Whereas, the study by Kumar and Mittal (2020) stated that every type of application, including email sorting and computer vision, uses machine-automated learning algorithms to carry out impossibly difficult tasks. The researcher presents the most successful content-based spam filtering techniques which include social engineering and phishing. The study focuses on spam filtering and its variations that are based on machine learning, specifically on a thorough examination of malicious attempts.

As reported by Ghosh and Jalal (2020), their study proposed Machine Learning algorithms, which include Random Forest Models, Support Vector Machines (SVM), and Naïve Bayes to detect spam and malware in the email. Researchers will first collect a dataset from the Kaggle dataset, then they will analyze, detect spam emails, and investigate it using the three algorithms. The accuracy report of Support Vector Machines (SVM) is 0.90%, Naïve Bayes is 0.93%, and the Random Forest is 0.97%. The result revealed that Random-Forest outperformed the other two algorithms.

In the previous research of Gibert et al. (2020), a modern leading study targets the creation and use of machine learning approaches for detecting malicious software because of their capacity to keep up with malware evolution. This study strives to provide a thorough and organized analysis of machine learning methods for detecting malware, with a focus on deep learning methods. The survey assists researchers in developing an awareness of the malware detection field as well as the latest advancements and research areas being investigated by the scientific community to address the issue.

Rashid et al. (2020) revealed that the SVM classifier has the highest result for accurately identifying phishing sites with a 95.66% rate. Since everyone heavily relies on the Internet, then every one of them is prone to malware attacks. The proposed system, when tested against various typical phishing datasets included in the University of California Irvine (UCI) collection, the recommended approach shows promising results. As a result, the suggested method is selected and utilized for machine learning-based phishing detection.

In the study of Shhadat et al. (2020) over the past ten years, dependence on technology has increased considerably. This prompts attackers to create new malware that can carry out their destructive tasks, which could also involve causing damage or acquiring crucial information. Malware detection is therefore a vital ingredient of system security, particularly including those of smart and portable devices. This study seeks to investigate the machine learning algorithm, specifically Random Forest used to identify unrecognizable malware.

In line with the study of El Kouari et al. (2020) due to the incredible advancements achieved in mobile environments, social networks, online banking, cloud and web technologies, and smart networks, cybersecurity is a sector that is expanding and needs a lot of attention. This paper will assess and enumerate the different works that utilize machine learning for network security. Machine learning is used to combat phishing websites and spam emails.

Based on Gibson et al., (2020) on their paper, email is a popular form of communication and social interaction among several people, and that method of communication is exploited by spammers for phishing or spreading malware. Gibson et al. (2020) introduced a technique for identifying spam emails through the implementation of machine learning models: Random Forest, Naïve Bayes, Support Vector Machine, Multi-Layer Perceptron, and Decision Tree. They applied this approach to seven distinct email datasets and combined it with pre-processing techniques and feature extraction. The best overall results were obtained with Multinomial Naïve Bayes with Genetic Algorithm.

Zhou and Pang (2019) revealed that the majority of these attack vectors were found inside email attachments, and they took advantage of flaws in Adobe and Office programs. Many of these attack examples use PDF-based exploits. In contrast to previous research on identifying pdf malware, the researchers proposed Expdf, a reliable detection solution for exploitable code-based machine learning. Expdf proves its supremacy in exploit detection by achieving accuracy rates of 95.54% and recall rates of 97.54%. In addition to that, it could identify specific exploit vulnerability types.

The focus of the study of Zhou and Pang (2019) was to determine whether the PDF has malware within or nothing using machine learning. This related study also could identify other vulnerabilities in PDF. Using machine learning, malware detection of newly evolved malware can be possible to detect, which is why the researcher of this new study utilized it for detection in email content.

In the study of Sahingoz et al. (2019), the tremendous development of technology made consumers switch from traditional retail to online shopping which is why the attackers set out a new technique to keep up with the new shift. Determining whether the website is real or phishing is a highly difficult task. That is why the researcher of this study proposed a real-time anti-phishing system. The system that is being proposed will utilize seven distinct algorithms for classification and features based on natural language processing. After the experiment and comparative analysis of all algorithms, the Random Forest algorithm with natural language processing yielded a 97.88% accuracy rate for ULR phishing detection.

METHODOLOGY

Theoretical Framework

This section uses the experimental method applied to the study that will be used for the ensembled classifier, specifically the one that yields the best results for improved performance in filtering spam and malicious links in email. Hence, the author selected two top classifiers based on the test result for email spam filtering which is Random Forest, Support Vector Machine taken from the experiment of five best-performing classifiers which are K-Nearest Neighbors, Naïve Bayes, Decision Tree, Random Forest, and Support Vector Machine that would support the study.

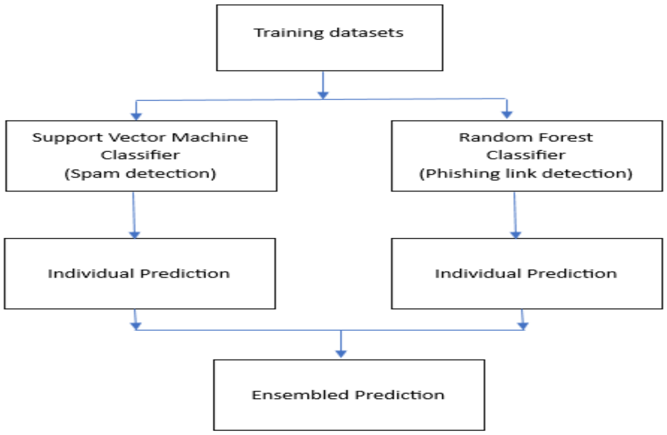
Training datasets
Support Vector Machine Classifier (Spam detection)
Random Forest Classifier (Phishing link detection)
Individual Prediction
Individual Prediction
Ensembled Prediction

Figure 1. Proposed Model Structure

Model Procedures

Data Collection

The sample test data was obtained from Kaggle of 5,573 datasets as it comprised a large collection of data of known spam and phishing links. It was used to evaluate the performance of the five chosen machine learning algorithms. The downloaded dataset was loaded in Jupyter Notebook using libraries such as Pandas.

Data Preprocessing

The data underwent a preprocessing phase to verify that the data had been cleaned and formatted correctly and was in a consistent structure. This step addressed any inconsistencies or issues that could potentially impact the performance of the selected machine-learning model.

Data Labeling

The labels assigned to the data were verified to ensure accuracy and consistency. Proper labeling of the data was important to provide the model with the necessary information to learn and make predictions. By doing so, the model could learn effectively with the data correctly representing the spam and phishing links.

Data Splitting

The data had been divided into separate training and testing datasets to evaluate the chosen model. It was done in a balanced and representative manner to ensure the reliable performance of the model by evaluating the unseen data properly.

Model Training

The machine learning models had been accurately trained and had effectively learned from the provided training data. Through this, the model would adjust and improve its ability to recognize patterns and make predictions based on the training data.

Model Evaluation

Using various evaluation metrics such as recall, precision, accuracy, F1 score, true positive rate, false positive rate, and confusion matrix, the model was able to accurately identify phishing emails within the testing dataset. These metrics provided insights into the model’s accuracy and effectiveness in detecting spam and phishing attempts.

Model Selection

Each model was verified for its capability to accurately detect spam and phishing links through the given dataset. The testing process produces the best models to be selected for the ensembled algorithm that supports the study.

Presentation of Formula

$$A = \frac{TP + TN}{TP + TN + FP + FN}$$

Equation 1. Accuracy Formula

Accuracy is a metric that measures the number of accurate predictions generated by a classifier on the tested data. To compute, the total number of true predictions, both true positives and true negatives, divided by the total number of predictions, which includes true and false positives and true and false negatives. The higher the accuracy result, the higher the quality of being true and correct of the classifier's prediction.

$$P = \frac{TP}{TP + FP}$$

Equation 2. Precision Formula

Precision measures the number of true predictions or exactness of the algorithm. The computation entails dividing the count of true positives by the sum of the total number of predicted positive instances. The higher the Precision result, the prediction of a classifier is more likely to be accurate.

$$R = \frac{TP}{TP + FN}$$

Equation 3. Recall Formula

Recall gauges a system's ability to detect all positive instances that the classifier detected from the entire set of positive samples. It is computed by dividing the number of true positives by the sum of true positives and false negatives. A high recall indicates that more positive samples have been detected.

$$F = \frac{2TP}{2TP + FP + FN}$$

Equation 4. F1 Score Formula

F1 score is a technique of evaluating the effectiveness of the system by summing up the predictive performance of the system by combining the two previously mentioned formulas — precision and recall. F1 score can fall within a certain range of values between 0 and 1, therefore the closer it is to 1, the better the system is. The higher the precision and recall, the higher the F1 score.

RESULTS

This section presented the experimental results obtained from training and testing various machine learning models using Kaggle public datasets. Metrics such as precision, recall, F1-score, and accuracy were used to assess the performance of each model (Tables 1-3). The strengths and weaknesses of each model were discussed, and their performance was compared to determine the most effective algorithm for spam and phishing detection (Figures 2-10).

Table 1. Summary of Test Results for Spam Detection

	Accuracy	Precision	Recall	F1-Score
Naive Bayes	0.88	0.54	0.86	0.66
SVM	0.97	1.00	0.81	0.90
KNN	0.89	1.00	0.20	0.34
Random Forest	0.97	1.00	0.76	0.86
Decision Tree	0.96	0.92	0.81	0.86

Table 2. Summary of Test Results for Phishing Detection

	Accuracy	Precision	Recall	F1-Score
Naive Bayes	0.82	0.77	1.00	0.87
SVM	0.86	0.86	0.91	0.88
KNN	0.78	0.80	0.84	0.82
Random Forest	0.87	0.88	0.89	0.88
Decision Tree	0.87	0.88	0.89	0.88

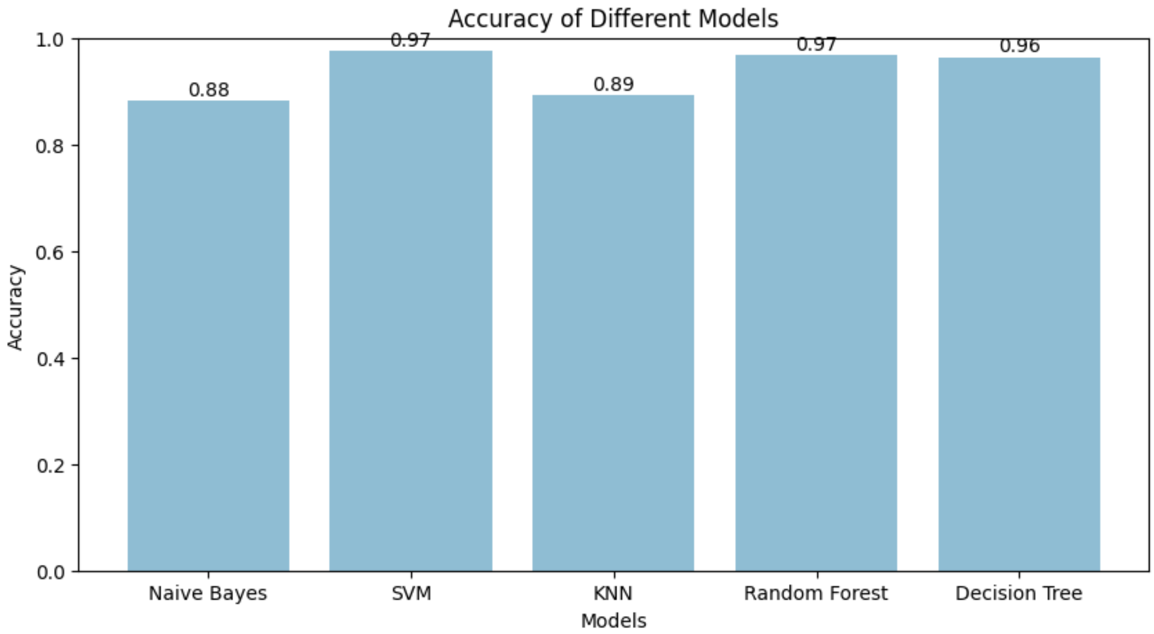

Figure 2. Bar Graph for Accuracy Comparison (Spam Detection)

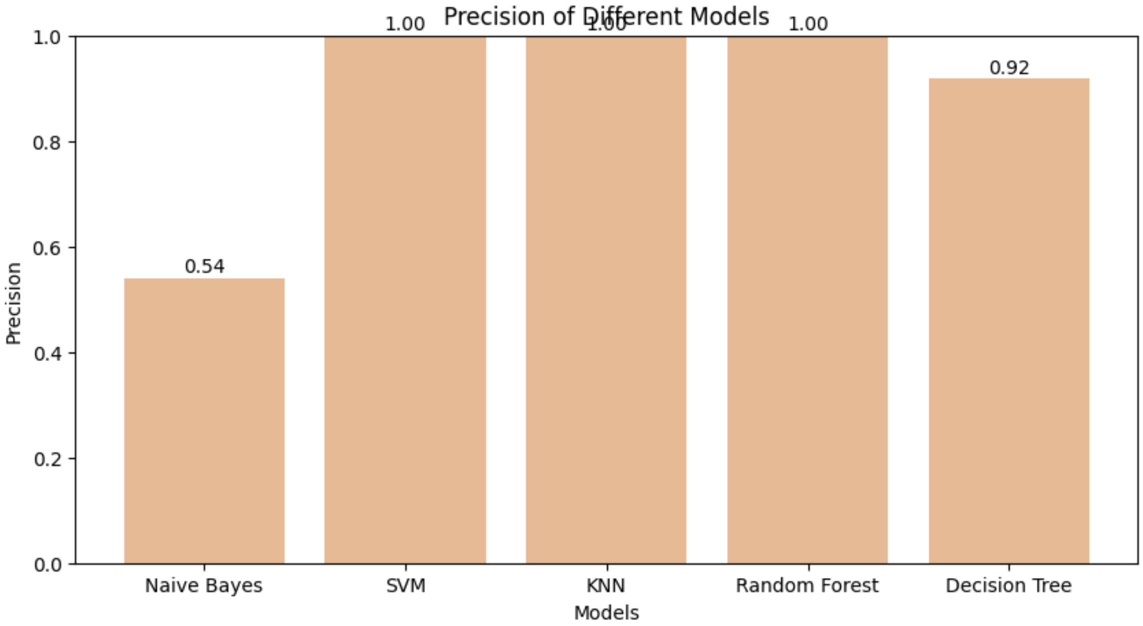

Figure 3. Bar Graph for Precision Comparison (Spam Detection)

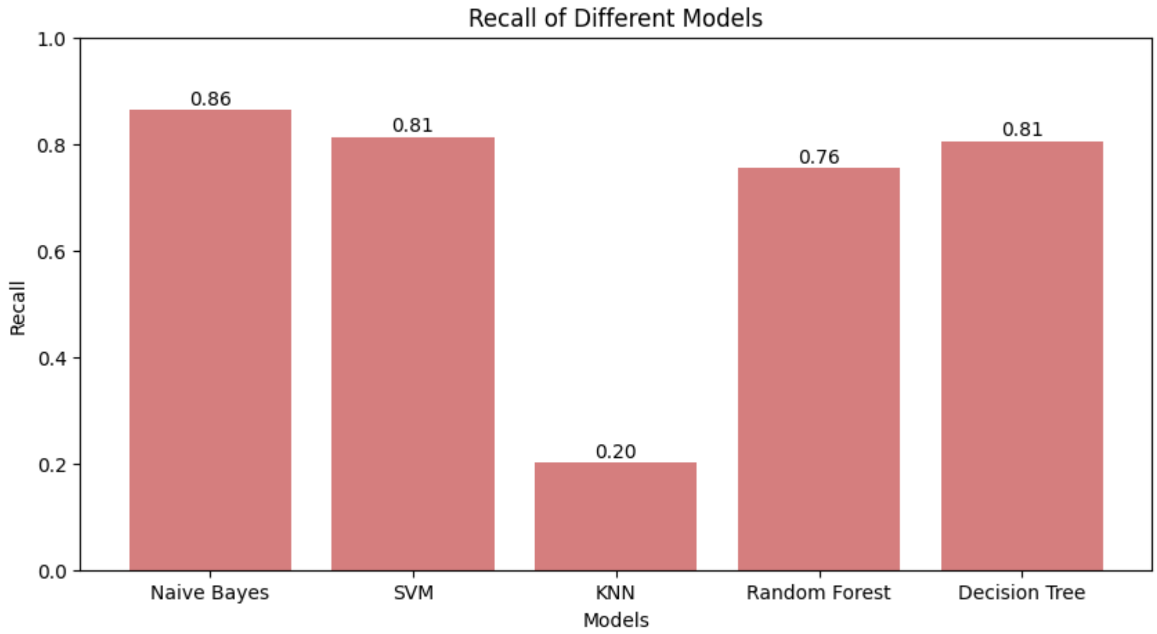

Figure 4. Bar Graph for Recall Comparison (Spam Detection)

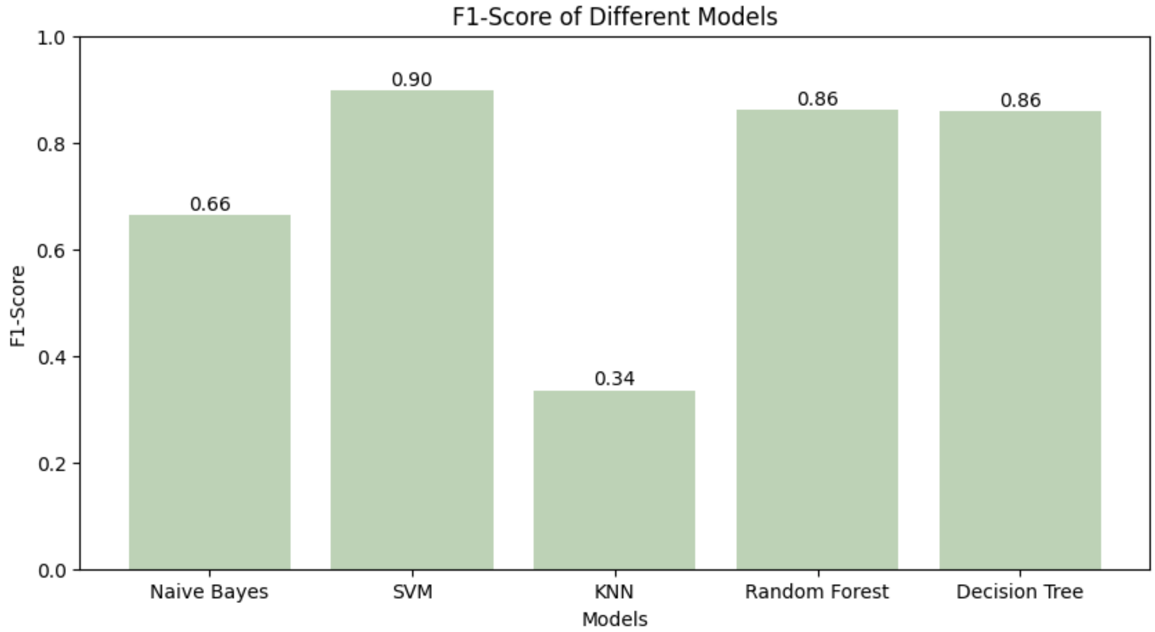

Figure 5. Bar Graph for F1-Score Comparison (Spam Detection)

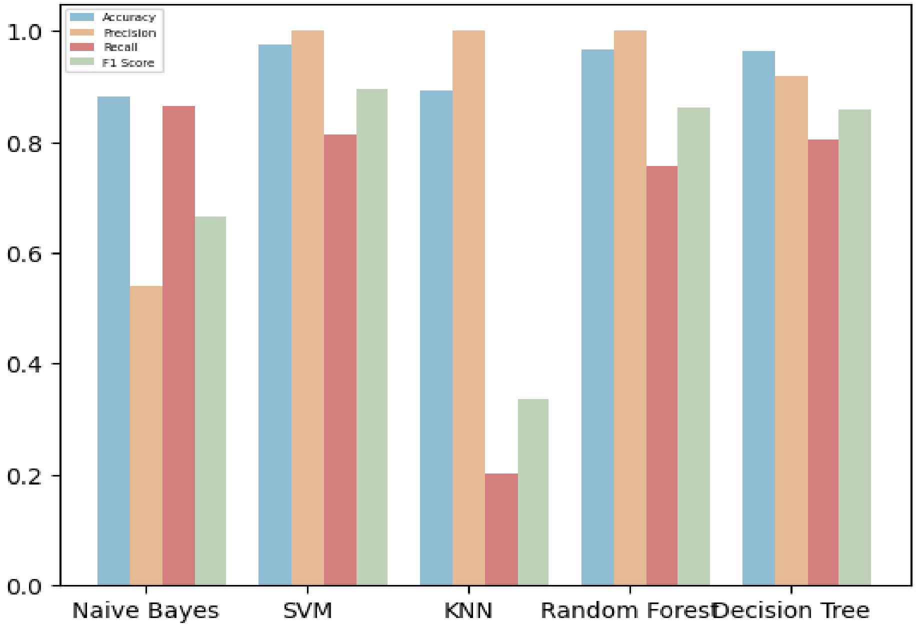

Figure 6. Overall Performance of Models (Spam Detection)

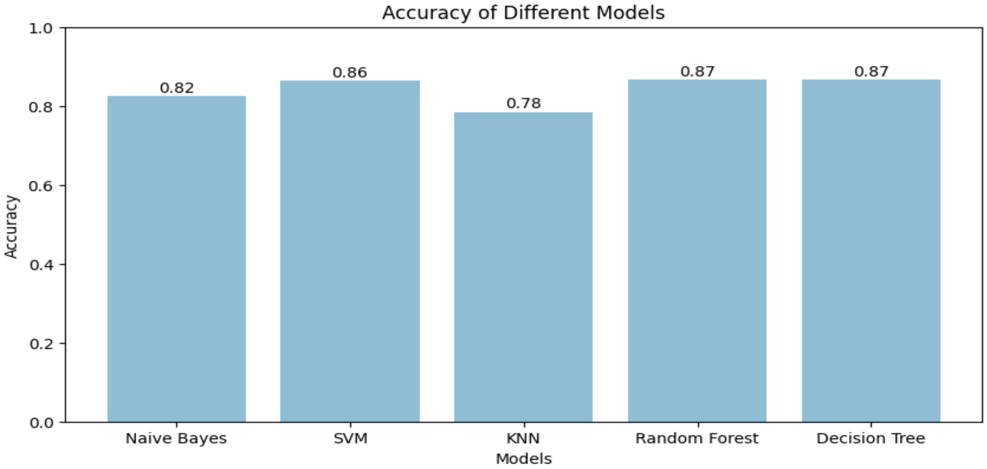

Figure 7. Bar Graph for Accuracy Comparison (Phishing Detection)

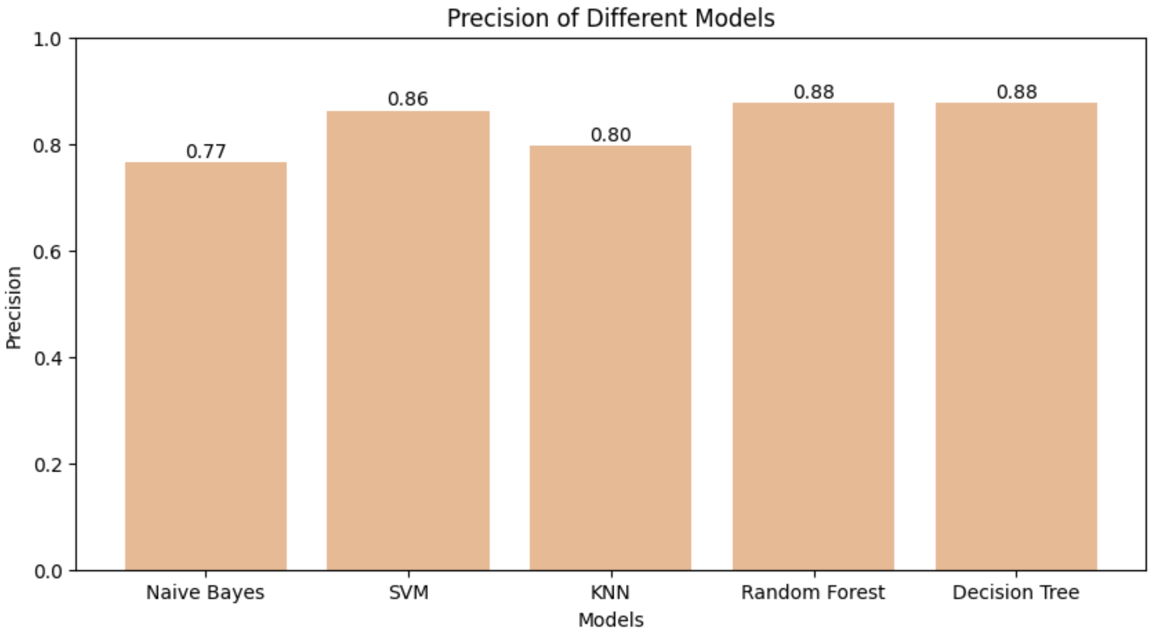

Figure 8. Bar Graph for Precision Comparison (Phishing Detection)

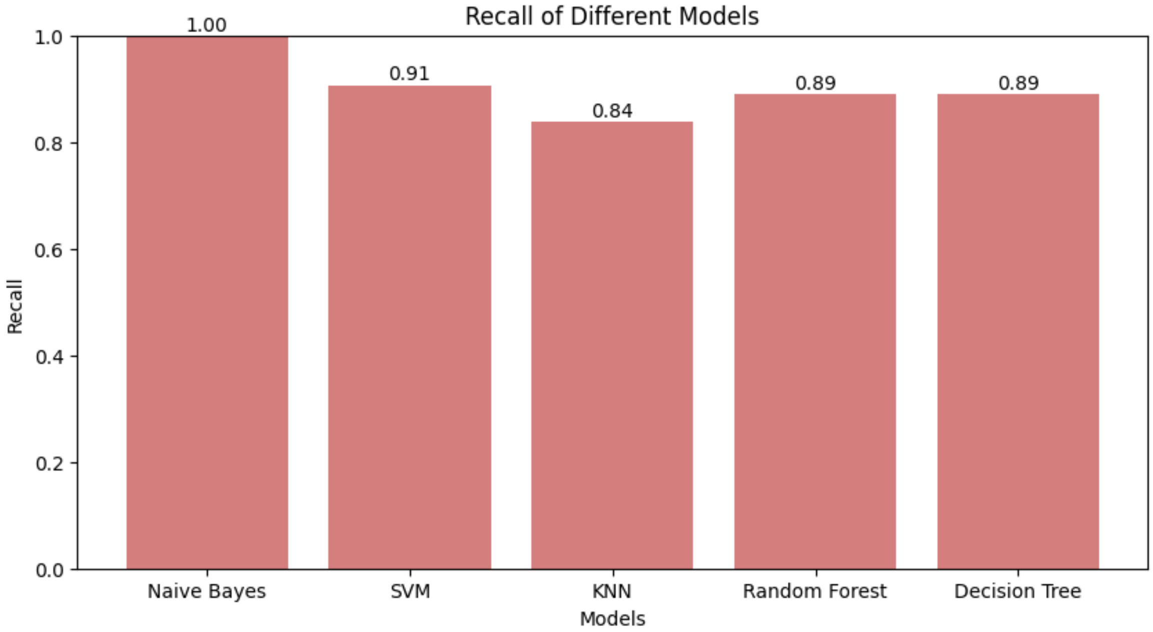

Figure 9. Bar Graph for Recall Comparison (Phishing Detection)

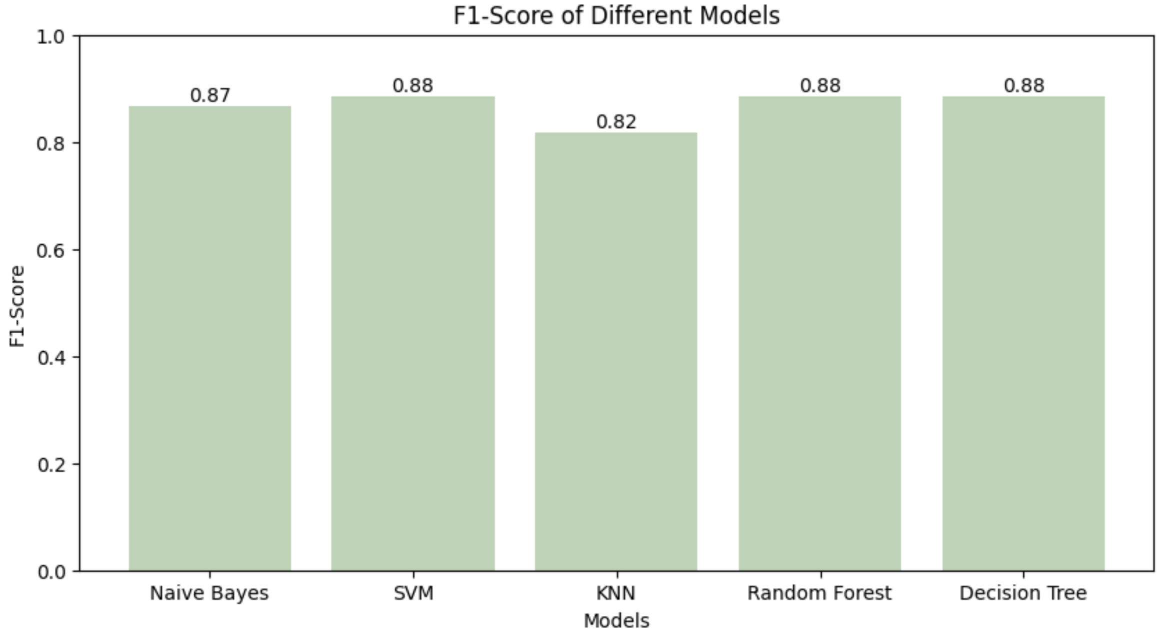

Figure 10. Bar Graph for F1-Score Comparison (Phishing Detection)

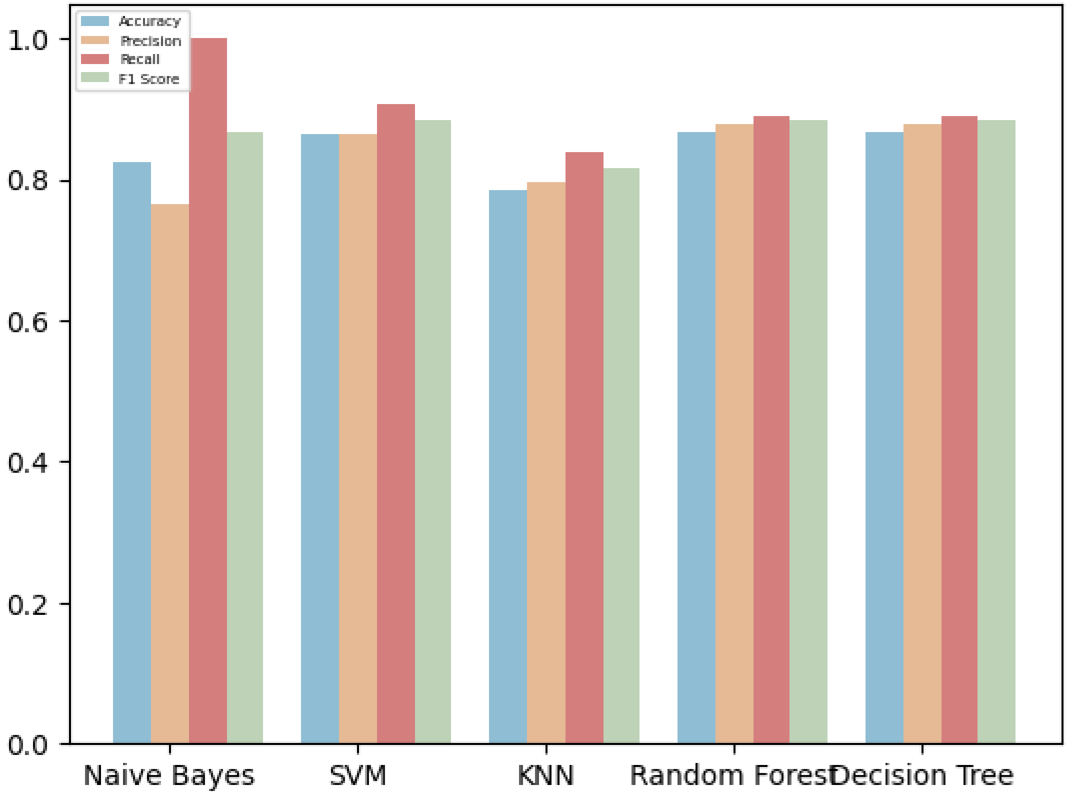

Figure 11. Overall Performance of Each Model (Phishing Detection)

Simulated Flow

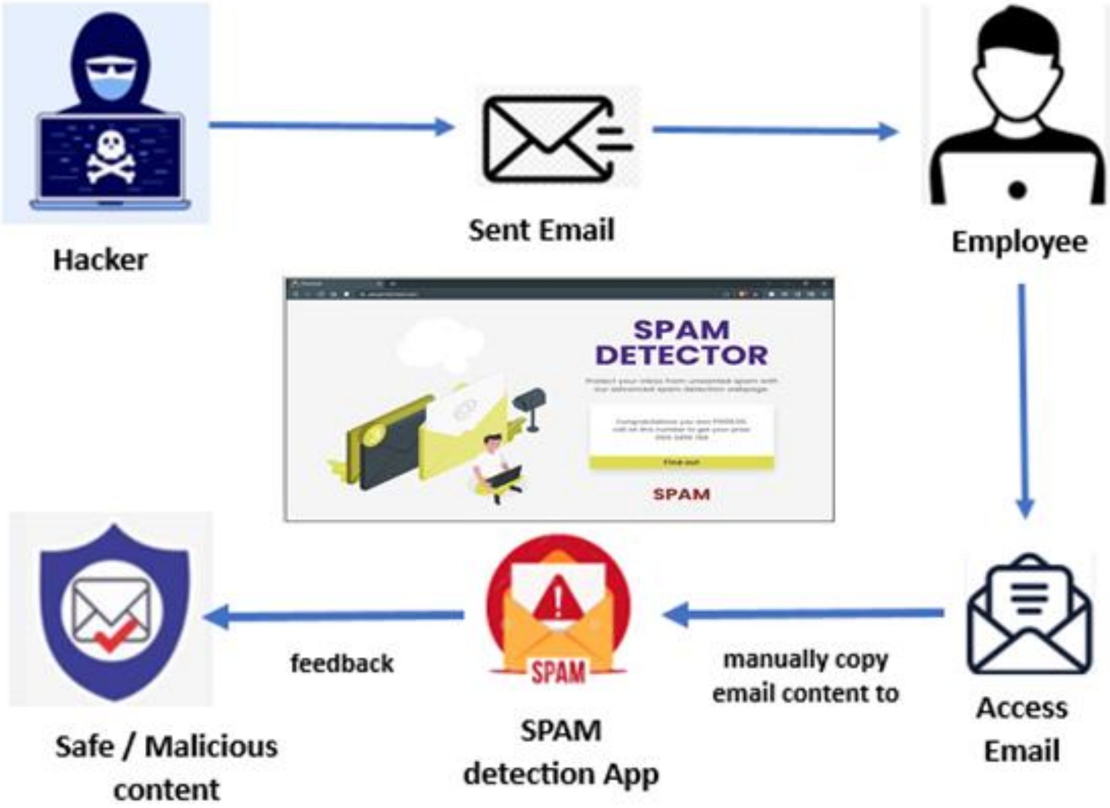

Figure 12. Spam Detection Web Page Structure

Figure 12 outlines the process of detecting spam emails in a workplace setting. The process begins with an unknown email sender sending an email to an employee's email application. The employee opens the email and views the email content. If the employee suspects that the email may be spam, they can copy and paste the email to the SPAM detection App: a spam detection website. The website then applies a Support Vector Machine (SVM) algorithm to identify if the email is spam or not. Depending on the result of the algorithm, the email is labeled as either safe email or spam email.

This figure helps highlight the importance of implementing spam detection measures in the workplace to prevent employees from falling victim to potentially harmful emails. By using a combination of human judgment and automated algorithms, organizations can reduce the risk of email threats and protect their sensitive information.

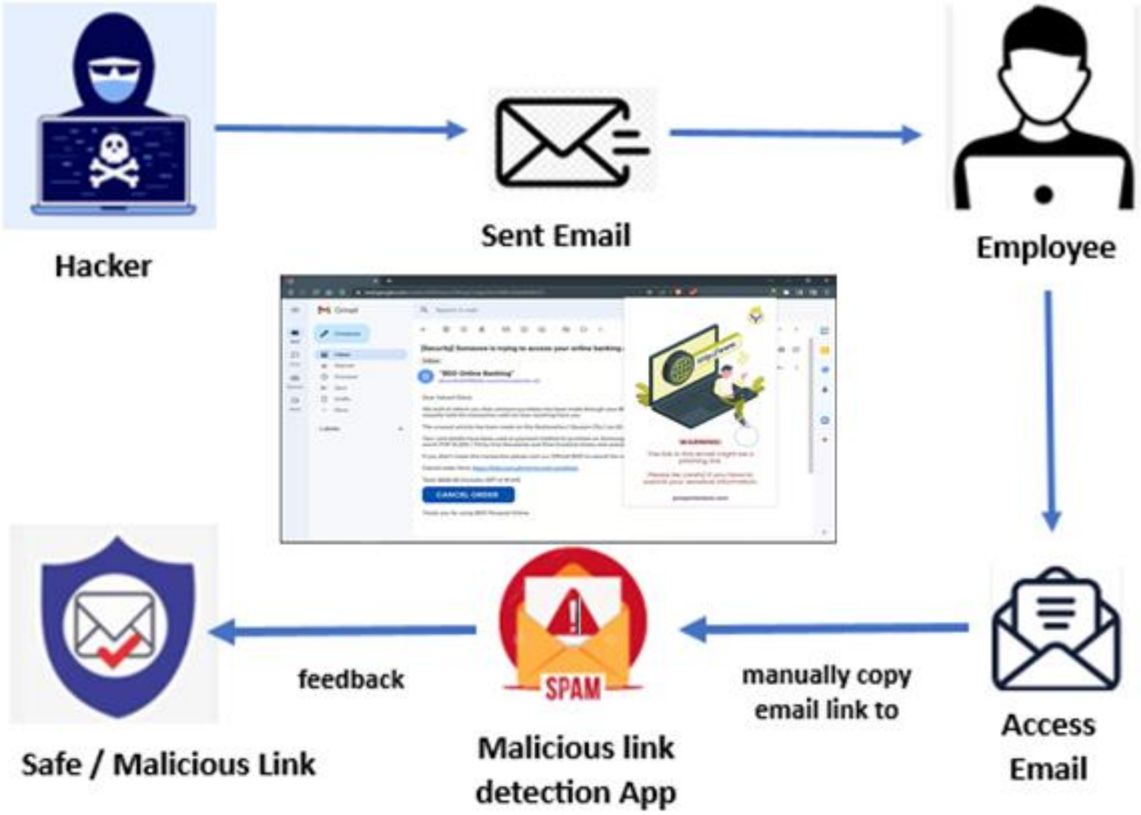

Figure 13. Browser Extension for Phishing Link Detection

Figure 13 depicts the process of a browser extension that uses the Random Forest machine learning algorithm to detect phishing links embedded in email content. When a user receives an email with malicious links, they can simply copy and paste the link into the extension. The browser extension detects if the link is suspicious since Random Forest is utilized and it has been trained on a large dataset of known phishing and non-phishing links. This enables the system to accurately identify potential threats and inform the user accordingly. If a link is classified as phishing, the browser extension would take steps to alert the user by displaying a warning message. On the other hand, if a link is classified as non-phishing, the extension will notify the user to safely click on the link and visit the website.

DISCUSSION

The experimental development of the ensemble algorithm was designed to validate the reliable detection of fraudulent emails that may have posed a security risk to email users as compared to the single best-selected algorithm from other studies. The system was capable of detecting spam and phishing emails by analyzing various features such as the embedded link and malicious content of the message. It uses two machine learning algorithms that garnered the highest accuracy to analyze these features and detect any suspicious patterns that may have indicated the presence of fraudulent email activity.

Table 1, shows that SVM and Random Forest algorithms have the highest accuracy and F1-score, with SVM having perfect precision and Random Forest having the highest precision and recall. The naive Bayes algorithm has the lowest precision and F1 score among the five algorithms but still performs reasonably well in accuracy and recall. The KNN algorithm has the highest precision, but its recall is significantly lower, leading to a low F1 score. The decision tree algorithm shows a balanced performance in terms of accuracy, precision, recall, and F1-score but is not as high as SVM and Random Forest.

Yet, Table 2 shows that among the five algorithms, Naive Bayes has the highest recall of 1.00, indicating that it correctly identifies all phishing instances in the test dataset. However, its precision is the lowest among all algorithms, meaning that it has a relatively high number of false positives. Random Forest and Decision Tree have the same scores for all evaluation metrics, indicating that they have similar performance in phishing detection. On the other hand, SVM has an accuracy of 0.86, indicating that it has the second-highest percentage of correct predictions. Its precision and F1-score are also high, indicating that it has a good balance between identifying phishing instances and minimizing false positives. KNN has the lowest accuracy of 0.78, but its precision and recall are both above 0.80, indicating that it has a relatively low number of false positives and false negatives.

While Figure 2, is a bar graph that compares the accuracy of various spam detection methods, including Naive Bayes, SVM, KNN, Random Forest, and Decision Tree. The graph displays the accuracy scores for each method, with Naive Bayes scoring 0.88, SVM having a score of 0.97, KNN achieving 0.89, Random Forest having a score of 0.97, and Decision Tree achieving an accuracy score of 0.96. The graph visually represents the performance of each algorithm and indicates that SVM and Random Forest have the highest accuracy, while Naive Bayes and KNN have lower accuracy. This comparison can be used to evaluate which method is most effective at detecting spam, based on its accuracy score.

Figure 3, presents a precision comparison of spam detection methods, with Naive Bayes achieving 0.54, SVM, KNN, and Random Forest achieving a precision of 1, and Decision Tree achieving 0.92. The precision metric measures the accuracy of a model's positive predictions such as the percentage of spam emails that were correctly identified as spam. The high precision values obtained by SVM, KNN, and Random Forest models in this comparison indicate that they are effective at identifying spam emails, making them promising options for spam detection applications.

While Figure 4, compares the recall scores of different machine-learning algorithms used for spam detection. The graph displays the recall scores for five algorithms: Naive Bayes (0.86), SVM (0.81), KNN (0.20), Random Forest (0.76), and

Decision Tree (0.81). The recall metric in spam detection is to measure the algorithms' ability to correctly identify all relevant spam emails, minimizing the chances of false negatives.

Also, Figure 5 compares the F1-score performance of different algorithms used for spam detection, with SVM (Support Vector Machine) having the highest f1-score of 0.90, followed by Random Forest and Decision Tree, both with an f1-score of 0.86. Naive Bayes had an f1-score of 0.66, while KNN (K-Nearest Neighbors) had the lowest score of 0.34. F1-score helps to determine the effectiveness of the algorithm in detecting spam messages accurately while minimizing false positives and false negatives.

Figure 6, shows that the best algorithm for spam detection would be SVM, as it has the highest accuracy, precision, and F1 score among the five algorithms tested. SVM achieved an accuracy of 0.97, indicating that it correctly classified 97% of the emails. It also achieved perfect precision, correctly classifying all spam emails, which is essential to avoid false positives. Its recall score of 0.81 indicates that it correctly identified 81% of the actual spam emails, and its F1 score of 0.90 is the highest among the five algorithms tested, indicating a good balance between precision and recall. Therefore, SVM is the most suitable algorithm for spam detection on a website, providing accurate and reliable results.

While Figure 7, shows that Random Forest and Decision Tree models achieved the highest accuracy of 0.87, followed by SVM with 0.86. Naive Bayes achieved an accuracy score of 0.82, while KNN achieved the lowest accuracy score of 0.78. This suggests that Random Forest and Decision Tree models are the most effective at accurately identifying phishing emails, with SVM also performing well in terms of accuracy. However, Naive Bayes and KNN may not be as accurate and reliable as the other models, indicating that they may be less suitable for real-world phishing detection applications where accuracy is crucial.

Also, Figure 8 shows that Naive Bayes has the lowest precision score of 0.77, meaning that out of all the links classified as phishing by the model, only 77% are phishing emails. SVM has the highest precision score of 0.86, indicating that it correctly identifies 86% of phishing emails. KNN has a precision score of 0.80, while both Random Forest and Decision Tree have a precision score of 0.88, suggesting that they are better at correctly identifying phishing links than the other models tested.

Based on Figure 9, we can see that Naive Bayes achieved perfect recall, correctly identifying all instances of phishing in the dataset. SVM had a recall score of 0.91, which means it identified 91% of phishing instances. Random Forest and Decision Tree models had the same recall score of 0.89, indicating that they correctly identified 89% of phishing instances in the dataset. KNN had a lower recall score of 0.84, which suggests that it missed some instances of phishing. Overall, the Naive Bayes model performed the best in terms of recall for phishing detection.

In Figure 10, it is evident that the SVM, Random Forest, and Decision Tree models outperform the other models in terms of the F1 score. They all achieved the same F1-scores of 0.88, indicating a good balance between precision and recall. Naive Bayes achieved a relatively high F1-score of 0.87, but its precision is comparatively lower than the other models. On the other hand, KNN has the lowest F1-score of 0.82 among all the models, indicating that it has a poorer balance between precision and recall compared to the other models.

Based on the results from Figure 11, Random Forest and Decision Tree models are the best algorithms for phishing detection. Both models achieved identical accuracy, precision, recall, and F1-scores of 0.87, 0.88, 0.89, and 0.88. These high scores indicate that these models are effective in detecting phishing emails. The high precision rate of these models ensures that false positives are minimized, while their high recall rate means that a significant proportion of actual phishing emails are identified. Although SVM also performed well, the Random Forest and Decision Tree models outperformed it in terms of precision and F1-score.

For best choice therefore, based on the results in the selection for the ensembled algorithm that best fits with spam detection would be Support Vector Machine (SVM) while the best algorithm for handling phishing detection (phishing links) would be Random Forest. These two algorithms were the top picked by the author to combine and work hand-in-hand as ensembled machine learning algorithms as recommended solutions to the gap of existing studies found in the literature review on email detection that fall short due to their approach based only on a best selected single algorithm that does all the task on email detections.

CONCLUSIONS AND RECOMMENDATIONS

The summary of test results shows that both algorithms achieved high accuracy rates in detecting spam and phishing links, with the SVM algorithm which achieves an accuracy of 0.97%, and the Random Forest algorithm which achieves an accuracy of 0.87%. The author also compared these algorithms with other approaches and found that SVM and Random Forest outclassed them in terms of accuracy, precision, recall, f1 score, true positive rate, and false positive rate.

The results suggest that ensembled machine learning algorithms can be effective in detecting spam and malicious links in email content as compared to other studies that depend the detection only on the best single selected algorithm and not an ensembled one. The high accuracy rates achieved by both models indicate that they can be used as reliable ensembled tools for email security.

It is highly recommended to embed the tested sample system and the like in several email providers for automatic detection and prevention of phishing emails from user access. The email system should allow users to detect spam without having to copy and paste the email content into a webpage. Also, disabling malicious links and detecting malicious email attachments (payloads) should be included to further the capabilities of this study.

THEORETICAL IMPLICATIONS

The study on ensembled algorithms in machine learning and deep learning if carefully selected will surely advance the accuracy detection of false positives or false negatives in email. This will lead to trust and worry-free email usage for everyone. The experimental development of a spam and phishing link detection system using machine learning algorithms has shown great promise in improving email security. The testing and sample system can be further improved and can provide a valuable tool for individuals and organizations to protect against spam and phishing attacks.

DECLARATIONS

Conflict of Interest

I hereby certify that this work was through my intellectual initiatives to the best of my knowledge and has not been published by others in any publication website or journal.

Informed Consent

I have read and understand the provided guidelines in this journal publication and had the opportunity to ask questions. I understand that my participation is voluntary and that I am fully knowledgeable about this publication and all its rules and regulations.

Ethics Approval

I declare adherence to the accepted ethical standards and my confidence in the originality of my research study.

REFERENCES

Akhtar, M. S., & Feng, T. (2022). Malware Analysis and Detection Using Machine Learning Algorithms. *Symmetry, 14(11),* 2304.

Alani, M. M., & Tawfik, H. (2022). PhishNot: a cloud-based machine-learning approach to phishing URL detection. *Computer Networks, 218,* 109407.

Ahmed, N., Amin, R., Aldabbas, H., Koundal, D., Alouffi, B., & Shah, T. (2022). Machine learning techniques for spam detection in email and IoT platforms: Analysis and research challenges. *Security and Communication Networks, 2022,* 1-19.

Alhogail, A., & Alsabih, A. (2021). Applying machine learning and natural language processing to detect phishing emails. *Computers & Security, 110,* 102414.

Alkhalil, Z., Hewage, C., Nawaf, L., & Khan, I. (2021). Phishing attacks: A recent comprehensive study and a new anatomy. *Frontiers in Computer Science, 3,* 563060.

Baig, A. (2021). Email Spam Detection using SVM. *International Journal for Research in Applied Science and Engineering Technology, 9(VII)*, 669-672. Retrieved from https://www.deepdyve.com/lp/unpaywall/email-spam-detection-using-svm-scXJ3cQ5HU?articleList=%2Fsearch%3Fquery%3Demail%2Bspam%2Bdetection%2Busing%26docNotFound%3Dtrue

Bawazeer, O., Helmy, T., & Al-Hadhrami, S. (2021, July). Malware detection using machine learning algorithms based on hardware performance counters: analysis and simulation. *In Journal of Physics: Conference Series* (Vol. 1962, No. 1, p. 012010). IOP Publishing. Retrieved from https://iopscience.iop.org/article/10.1088/1742-6596/1962/1/012010

Trivedi, H., & Broadhurst, R. (2020). Malware in spam email: Risks and trends in the Australian Spam Intelligence Database. *Trends and Issues in Crime and Criminal Justice [electronic resource], (603),* 1-18. Retrieved from https://search.informit.org/doi/abs/10.3316/agispt.20201215041188

Ding, X., Liu, B., Jiang, Z., Wang, Q., & Xin, L. (2021, May). Spear phishing email detection based on machine learning. *In 2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD)* (pp. 354-359). IEEE,

Ghosh, S., & Jalal, S. (2020). Email Spam and Malware Detection using Machine Learning. *International Research Journal of Modernization in Engineering Technology and Science, 2(9),* 1401-1404.

Gibson, S., Issac, B., Zhang, L., & Jacob, S. M. (2020). Detecting spam email with machine learning optimized with bio-inspired metaheuristic algorithms. *IEEE Access, 8,* 187914-187932.

Gibert, D., Mateu, C., & Planes, J. (2020). The rise of machine learning for detection and classification of malware: Research developments, trends, and challenges. *Journal of Network and Computer Applications, 153,* 102526.

Haiba, S., & Mazri, T. (2021, April). Build a malware detection software for IoT network using machine learning. *In Proceedings of the 4th International Conference on Networking, Information Systems & Security* (pp. 1-8).

Karim, A., Azam, S., Shanmugam, B., Kannoorpatti, K., & Alazab, M. (2019). A comprehensive survey for intelligent spam email detection. *IEEE Access, 7,* 168261-168295.

Khan, R., Kumar, N., Handa, A., & Shukla, S. K. (2021). Malware Detection in Word Documents Using Machine Learning. In *Advances in Cyber Security: Second International Conference, ACeS 2020, Penang, Malaysia, December 8-9, 2020, Revised Selected Papers* 2 (pp. 325-339). Springer Singapore. Retrieved from https://link.springer.com/chapter/10.1007/978-981-33-6835-4_22

El Kouari, O., Benaboud, H., & Lazaar, S. (2020, March). Using machine learning to deal with Phishing and Spam Detection: An overview. In *Proceedings of the 3rd International Conference on Networking, Information Systems & Security* (pp. 1-7).

Kumar, S., & Mittal, S. K. (2020). Email spam and malware filtering using machine learning and its applications. *Performance Management, 2*(2), 25-32.

Lee, C., & Lee, K. (2022). Impact Analysis of Resilience Against Malicious Code Attacks via Emails. *Computers, Materials & Continua*, *72*(3). Retrieved from https://cdn.techscience.cn/ueditor/files/cmc/TSP_CMC-72-3/TSP_CMC_25310/TSP_CMC_25310.pdf

Mughaid, A., AlZu'bi, S., Hnaif, A., Taamneh, S., Alnajjar, A., & Elsoud, E. A. (2022). An intelligent cyber security phishing detection system using deep learning techniques. *Cluster Computing, 25*(6), 3819-3828.

Muralidharan, T., & Nissim, N. (2023). Improving malicious email detection through novel designated deep-learning architectures utilizing entire email. *Neural Networks, 157,* 257-279.

Ojewumi, T. O., Ogunleye, G. O., Oguntunde, B. O., Folorunsho, O., Fashoto, S. G., & Ogbu, N. J. S. A. (2022). Performance evaluation of machine learning tools for detection of phishing attacks on web pages. *Scientific African, 16,* e01165.

Phomkeona, S., & Okamura, K. (2020). Zero-day malicious email investigation and detection using features with a deep-learning approach. *Journal of Information Processing, 28,* 222-229.

Ouyang, T. (2022). *On Mitigating Email Spam and Phishing URLs* (Doctoral dissertation). Case Western Reserve University, Ohio, USA. Retrieved from https://www.proquest.com/openview/afc08ce0f36a321d8d324ac9581875f1/1?pq-origsite=gscholar&cbl=18750&diss=y

Quang, D. N., Selamat, A., & Krejcar, O. (2021). Recent research on phishing detection through machine learning algorithm. In *Advances and Trends in Artificial Intelligence. Artificial Intelligence Practices: 34th International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, IEA/AIE 2021, Kuala Lumpur, Malaysia, July 26–29, 2021, Proceedings, Part I 34* (pp. 495-508). Springer International Publishing. Retrieved from https://link.springer.com/chapter/10.1007/978-3-030-79457-6_42

Rashid, J., Mahmood, T., Nisar, M. W., & Nazir, T. (2020, November). Phishing detection using machine learning technique. *In 2020 first international conference of smart systems and emerging technologies (SMARTTECH)* (pp. 43-46). IEEE.

Sahingoz, O. K., Buber, E., Demir, O., & Diri, B. (2019). Machine learning-based phishing detection from URLs. *Expert Systems with Applications, 117,* 345-357.

Schott, M. (2019). Random forest algorithm for machine learning. *medium. com.* Retrieved from https://medium. com/capital-one-tech/random-forest-algorithm-for-machine-learning-c4b2c8cc9feb (Erişim 4 Ocak 2021).

Shhadat, I., Hayajneh, A., & Al-Sharif, Z. A. (2020). The use of machine learning techniques to advance the detection and classification of unknown malware. *Procedia Computer Science, 170,* 917-922.

Siddique, Z. B., Khan, M. A., Din, I. U., Almogren, A., Mohiuddin, I., & Nazir, S. (2021). *Machine learning-based detection of spam emails. Scientific Programming, 2021,* 1-11.

Singh, J., & Singh, J. (2021). A survey on machine learning-based malware detection in executable files. *Journal of Systems Architecture, 112*, 101861.

Sohail, B. (2021). Macro-Based Malware Detection System. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *12*(3), 5776-5787.

Swetha, M. S., & Sarraf, G. (2019, May). Spam email and malware elimination employing various classification techniques. In *2019 4th International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT)* (pp. 140-145). IEEE.

Wei, W., Ke, Q., Nowak, J., Korytkowski, M., Scherer, R., & Woźniak, M. (2020). Accurate and fast URL phishing detector: a convolutional neural network approach. *Computer Networks, 178,* 107275.

Wu, P., & Guo, H. (2022, December). Holmes: An efficient and lightweight semantic-based anomalous email detector. In *2022 IEEE International Conference on Trust, Security, and Privacy in Computing and Communications (TrustCom)* (pp. 1360-1367). IEEE.

Zhang, J., Li, W., Gong, L., Gu, Z., & Wu, J. (2019, December). Targeted malicious email detection using hypervisor-based dynamic analysis and ensemble learning. In *2019 IEEE Global Communications Conference (GLOBECOM)* (pp. 1-6). IEEE.

Zhou, X., & Pang, J. (2019). Expdf: Exploits detection system based on machine learning. *International Journal of Computational Intelligence Systems*, *12*(2), 1019-1028.

Author's Biography

Erwin E. Guerra was a graduate of Bachelor of Science in Computer Engineering at Technological Institute of the Philippines (Manila Campus), also a graduate of Master in Information Technology at Polytechnic University of the Philippines (Sta. Mesa Campus), and currently finishing his Doctor of Information Technology degree (dissertation phase) in Technological Institute of the Philippines (Quezon City Campus). His field of interests include Robotics, Cloud Computing, Internet Security, Cybersecurity, Data Mining, Machine Learning, and Deep Learning.