JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Email Phishing Detection Using Machine Learning

Iqra Fayaz[1] and Zahoor Ahmad Nagar[1]

[1]IT, Central University Of Kashmir, India

Abstract

Email phishing continues to pose a major threat to cybersecurity, re sulting in considerable financial losses and data breaches worldwide. This dissertation offers a thorough investigation into creating a system based on machine learning for detecting phishing emails. The main aim of this research is to utilize machine learning techniques to improve the accu racy and efficiency of phishing detection. Various algorithms, such as Support Vector Machine (SVM), Decision Tree, Naive Bayes, Random Forest, and Logistic Regression, are utilized to classify emails as either phishing or legitimate. Hyperparameters are fine-tuned to enhance the accuracy of these algorithms, and regularization methods are employed to address overfitting issues.The performance of these models is assessed us ing metrics such as accuracy, precision, recall, and F1-score. The findings reveal that the Random Forest algorithm, with optimized hyperparame ters, achieves the highest detection accuracy, significantly outperforming traditional methods. This study highlights the potential of machine learn ing in enhancing email security and provides a solid framework for future research in phishing detection. The results emphasize the necessity for continuous advancements in machine learning to protect against evolving cyber threats.

Keywords: Email phishing, Machine learning, Cybersecurity, Phishing de tection, Support Vector Machine (SVM), Decision Tree, Naive Bayes, Random Forest, Logistic Regression, Hyperparameters, Regularization methods, Detection accuracy, Precision, Recall, F1-score, Email security, Cyber threats.

1. Introduction

Electronic mail, or email, stands as one of the most essential features of the Internet, alongside the web. It facilitates the sending and receiving of messages between individuals with email addresses globally. Operating through multiple protocols within the TCP/IP suite, email has long been a crucial communication tool, providing nearly instantaneous connectivity to any location with internet access. However, phishing emails pose a significant threat within this commu nication medium. These emails are a type of spam specifically designed to trick recipients into

revealing personal information, such as passwords, credit card numbers, and bank account details. Falling victim to such scams can lead to f inancial fraud and identity theft. In some cases, these emails are meticulously crafted to target specific individuals to extract sensitive information about their colleagues or organization, known as spear-phishing.

1.1 Types of Email Phishing Attacks

1. Deceptive Phishing: This is the most common form of phishing, where attackers impersonate a legitimate company or individual to steal personal information. These emails often create a sense of urgency to provoke an immediate response.

2. Spear Phishing: Unlike deceptive phishing, spear phishing targets specific individuals or organizations. Attackers personalize the email content using information gathered from social media and other sources to increase the likelihood of success.

3. Whaling: This type of phishing targets high-profile individuals within an organization, such as executives or senior management. The emails are often crafted to address business-related issues, making them appear more credible.

4. Clone Phishing: In this attack, a legitimate and previously delivered email containing an attachment or link is cloned, and its content is replaced with malicious content. The attacker sends the modified email from an address resembling the original sender, tricking the recipient into believing it is genuine.

5. Pharming: Pharming involves redirecting users from a legitimate website to a fraudulent one. Although not strictly an email attack, phishing emails often contain links that lead to pharming websites designed to harvest login credentials and other sensitive information.

6. CEO Fraud (Business Email Compromise): This attack targets businesses working with foreign suppliers and/or businesses that regularly perform wire transfer payments. The attacker poses as the CEO or another high level executive and requests an urgent wire transfer to a seemingly legiti mate account.

1.2 Importance of the research

This research examines current spam detection mechanisms and identifies the shortcomings of existing email filtering strategies. It focuses on detecting phish ing emails by analyzing the email text to determine whether it is legitimate or a phishing attempt. The research aims to develop an effective machine learning based mechanism for detecting, filtering, and classifying phishing emails within organizations. Specifically, the study utilizes machine learning algorithms to enhance the accuracy and reliability of phishing email detection. By applying these machine learning techniques, this research seeks to provide a robust solu tion to the growing problem of phishing emails, thereby improving the security and integrity of email communications within organizations. The methodologies and findings presented in this thesis are expected to make a significant contri bution to the ongoing efforts to combat phishing attacks and protect sensitive information from being compromised. To enhance the accuracy of phishing email detection

systems and address the shortcomings of current spam detec tion methods, it is crucial to study these issues. Attackers constantly adapt new techniques, causing the content of phishing emails and malicious attachments to vary over time. The objective of our research was to develop an effective algorithm for phishing detection that can identify, prevent, and protect users from phishing emails containing malicious content. This algorithm aims to re duce the number of successful phishing attacks targeted at users. Furthermore, our research categorized various types of phishing and validated the algorithm's accuracy. The resulting solution is beneficial for both companies and employees seeking protection from malicious email links and phishing attacks.

2. Literature Survey

This section reviews several key research papers that focus on email phishing detection using machine learning algorithms. Keerthika J et al. [1] explore the application of various machine learning algorithms for classifying emails as either phishing or legitimate. Their methodology includes essential steps such as data pre-processing, feature extraction, and the implementation of different ML algo rithms. The pre-processing phase involves tokenization, removal of stop words, and handling missing values. The algorithms evaluated include Naive Bayes (NB), Logistic Regression (LR), Random Forest (RF), Decision Tree (DT), and Support Vector Machine (SVM). Their findings indicate that SVM and Naive Bayes algorithms exhibit superior performance in predicting phishing emails. They also suggest that tuning the hyperparameters of these algorithms can fur ther enhance their accuracy. The study proposes future research to integrate optimization algorithms with Naive Bayes and explore additional features to improve detection rates. The International Journal of Interactive Mobile Technologies (iJIM) [2] investi gates the detection of phishing attacks through the evaluation of various machine learning classifiers on three distinct datasets, both before and after feature selection using the Information Gain (IG) algorithm. The datasets include the UCI Phishing Websites dataset and two others with varying feature sets and labels. The study concludes that Random Forest consistently achieves the highest accu racy across all datasets, followed by J48, Artificial Neural Network (ANN), and K-Nearest Neighbors (KNN). The Decision Stump classifier, however, shows the lowest accuracy. The use of IG for feature selection enhances classifier perfor mance by reducing irrelevant features. Milda Tubyte et al. [3] analyze the effectiveness of different machine learning algorithms in classifying URLs as either phishing or legitimate. Their study addresses the URL Boolean classification problem, employing machine learn ing methods such as SVM, Random Forest, DT, Linear Discriminant Analysis (LDA), and LR. Utilizing two distinct datasets, they conduct a comparative analysis of these algorithms. The research aims to improve phishing detection by identifying malicious URLs in emails, which is a prevalent cybersecurity is sue. Fenny Zalavadia et al. [4] delve into the efficacy of Natural Language Pro cessing (NLP) and Deep Learning models for detecting phishing emails. Their proposed system, Phishector, utilizes NLP techniques to analyze email content and employs various Machine Learning and Deep Learning algorithms to clas sify emails as phishing or legitimate. The study uses the SpamAssassin and Ham-Spam datasets for training and testing. Results show that Random Forest and Extra Trees classifiers perform exceptionally well, with Neural Networks also yielding high accuracy under specific configurations. Despite these promising results, the study acknowledges limitations such as dataset dependency and potential overfitting issues with certain models. Tushaar et al. [5] focus on applying machine learning to

detect and filter spam and phishing emails, collectively referred to as Unsolicited Bulk Emails (UBEs). The study emphasizes the necessity for robust UBE filters due to the increasing volume and sophistication of these emails. They discuss feature extraction from email content and behavior, selection of discriminative features, and evaluation of various machine learning algorithms. Their findings suggest that machine learning models, particularly a combination of several state-of-the-art algorithms, can classify UBEs with high accuracy, achieving up to 99%. However, the paper highlights challenges such as the lack of specific dataset details and potential high computation time. Niranjan A et al. [6] present an approach to efficiently detect phishing by com bining feature selection methods with ensemble classifiers. The authors propose three feature selection algorithms: Mean of Mean of Ranks (MMR), Mean of Standard Deviation of Ranks (MSDR), and Feature Selection using Frequency (FSF). These methods aim to reduce the dataset's dimensionality while retain ing the most significant features. The selected features are then used in an ensemble model combining Random Committee and Random Forest classifiers with a StackingC technique (ERCRFS) to improve phishing detection accuracy. Meenu et al. [7] focus on developing a spam filter using various machine learn ing techniques to detect phishing emails. The study compares the predictive accuracy, F1 score, precision, and recall of several machine learning methods including Logistic Regression (LR), SVM, DT, and Neural Networks (NNet). The research aims to enhance phishing detection by improving the logistic re gression technique with feature selection methods.

3. Dataset Description and Proposed System

3.1 Unpacking the Dataset

In this project, multiple datasets were utilized to enhance the accuracy and reliability of phishing email detection using machine learning techniques. De tailed descriptions of the datasets are provided below:

• Phishing Email Dataset : The phishing email.csv dataset from Kaggle con tains 18,650 rows and 3 columns, with each row representing a single email, totaling 55,950 entries (18,650 rows * 3 columns). These columns usually encompass features like email content, metadata, and labels indicating if the email is phishing or legitimate. This setup enables thorough analysis and modeling, making it ideal for tasks related to email classification and phishing detection.

• Spam_Ham Dataset :The spamham.csv dataset from Kaggle contains 5,171 rows and 4 columns, resulting in a total of 20,684 entries (5,171 rows * 4 columns). Each row represents a single email record, with columns capturing various features related to the emails, such as email text, metadata, and labels indicating whether the email is classified as spam or ham. This structured dataset is suitable for analytical tasks like spam detection and classification, providing a robust basis for model training and evaluation.

3.2 Proposed System

Machine Learning in Email Phishing Detection Email phishing is a prevalent and damaging form of cyber attack. Machine learning offers a robust frame work for detecting phishing emails by analyzing various features and patterns within email content, metadata, and sender behavior. The application of ML in phishing detection involves several key steps:

1. Data Collection and Preprocessing:

• Data Collection: Gather a large dataset of emails, including both phishing and legitimate emails, for training and testing the machine learning models.

• Preprocessing: Clean and preprocess the data by removing noise, handling missing values, and converting the data into a suitable format for analysis. This step may include tokenizing email content, extracting features, and normalizing data.

2. Feature Extraction:

• Content Features: Extract features from the email body, such as the presence of certain keywords, the frequency of specific terms, and the overall tone of the message.

• Metadata Features: Analyze email metadata, including the sender's email address, domain reputation, and email headers.

• Behavioral Features: Examine patterns in sender behavior, such as the volume and frequency of emails sent, and any anomalies in sending patterns.

3. Model Training

• Supervised Learning: Use labeled datasets to train machine learning models. The model learns to differentiate between phishing and legitimate emails based on the extracted features.

• Evaluation: Assess the model's performance using metrics such as accu racy, precision, recall, and F1 score. Cross-validation techniques ensure that the model generalizes well to unseen data.

4. Deployment and Monitoring

• Real-Time Detection: Implement the trained model in an email filtering system to analyze incoming emails in real-time. Suspicious emails can be f lagged, quarantined, or blocked.

• Continuous Monitoring: Regularly monitor the model's performance and update it with new data to maintain its effectiveness against evolving phishing tactics.

3.2.1 Machine Learning Algorithms

1. Logistic Regression:

• Overview: Logistic Regression is used for binary classification tasks, predicting the probability that an instance belongs to a particular class.
• How It Works: Uses the logistic function to model the probability of the default class. The model outputs a probability value between 0 and 1, and a threshold (e.g., 0.5) is used to classify the email as phishing or legitimate.

2. Support Vector Machine (SVM):

• Overview: SVM is a powerful supervised learning algorithm for classification tasks, aiming to find the optimal hyperplane that separates classes with the maximum margin.
• How It Works: Constructs a hyperplane in a high-dimensional space to separate data points of different classes. For non-linearly separable data, SVM can use kernel functions to transform the data into a higher dimensional space where linear separation is possible.

3. Decision Tree:

• Overview: A flowchart-like model used for both classification and regression tasks, splitting the data into subsets based on input feature values.
• How It Works: Consists of nodes (decisions), branches (outcomes), and leaves (final decisions). Uses criteria such as Gini impurity or information gain to decide the best feature and threshold to split the data at each node. Pruning is used to avoid overfitting.

4. Naive Bayes:
• Overview: A probabilistic classifier based on Bayes' theorem with the assumption of feature independence.
• How It Works: Uses Bayes' theorem to compute the posterior probability of a class given the features. Despite the unrealistic assumption of feature independence, it is often effective.

5. Random Forest:
• Overview: An ensemble learning method that combines multiple decision trees to improve classification accuracy and robustness.
• How It Works: Constructs multiple decision trees during training, each built using a random subset of features and data samples. The final pre diction is made by aggregating the predictions of all trees.

Evaluation Metrics

1. Correlation Matrix:
• Overview: A table showing the correlation coefficients between variables, with each cell indicating the correlation between two variables.

2. Accuracy:

• Overview: The ratio of correctly predicted instances to the total instances, a basic metric for evaluating classification performance.

4. Evaluation of Results

Performance Evaluation of Classification Algorithms on the Spam _Ham Dataset

In the analysis of the Spam Ham dataset, which includes both spam and legitimate emails, we employed five different classification algorithms: Logistic Regression, Decision Tree, Support Vector Machine (SVM), Random Forest, and Naive Bayes. The evaluation used default parameters without hyperparameter tuning. Below are the performance metrics for each algorithm:

1.Logistic Regression

• Accuracy: 98%

• Logistic Regression achieved the highest accuracy, indicating effective dif- ferentiation between spam and legitimate emails.

2.Naive Bayes

• Accuracy: 98%

• Naive Bayes also achieved high accuracy, demonstrating its strength in probabilistic spam classification.

3.Support Vector Machine (SVM)

• Accuracy: 97%

• SVM showed robust performance, slightly lower than Logistic Regression and Naive Bayes.

4.Random Forest

• Accuracy: 97%

• Random Forest's ensemble approach yielded solid performance, comparable to SVM.

5.Decision tree

• Accuracy: 95%

• Decision Tree had the lowest accuracy, possibly due to overfitting or its simplistic nature in handling complex features.

Summary: Logistic Regression and Naive Bayes were the most accurate al- gorithms for the Spam Ham dataset, each achieving 98% accuracy. Decision Tree, while interpretable, had the lowest accuracy at 95%, suggesting that more advanced models might be more suitable for this task.

Dataset	Category	Algorithm	Precision	Recall	F1_Score	Accuracy
Spam ham dataset	Spam\ham	Logistic regression	0.99	0.99	0.99	98%
			0.97	0.97	0.97	
		Decision tree	0.98	0.96	0.97	95%
			0.90	0.94	0.92	
		Random forest	0.97	0.99	0.98	97%
			0.97	0.92	0.95	
		Naïve bayes	0.98	0.99	0.98	98%
			0.97	0.95	0.96	
		SVM	0.98	0.97	0.98	97%
			0.94	0.95	0.94	

Fig 1: Results of Spam_Ham dataset with default parameters

Impact of Hyperparameter Tuning on Algorithm Performance in case of Spam _Ham Dataset

The performance of machine learning algorithms on the Spam _Ham dataset was further evaluated with hyperparameter tuning:

1.Logistic Regression

• Accuracy with Default Parameters: 98%

• Accuracy with Hyperparameter Tuning: 98.26%

• Hyperparameter tuning led to a slight improvement.

2.Decision Tree

• Accuracy with Default Parameters: 95%

• Accuracy with Hyperparameter Tuning: 94.29%

• A decrease in accuracy, indicating potential overfitting.

3. **Random Forest**

• Accuracy with Default Parameters: 97%

• Accuracy with Hyperparameter Tuning: 97.48%

• Improvement in accuracy due to effective optimization

4.Naive Bayes

• Accuracy with Default Parameters: 98%

• Accuracy with Hyperparameter Tuning: 97.77%

• A slight decrease in accuracy.

5.Support Vector Machine (SVM)

• Accuracy with Default Parameters: 97%

• Accuracy with Hyperparameter Tuning: 97.39%

• Accuracy improved with tuning.

Summary:

• Increased Accuracy: Logistic Regression (98% to 98.26%), Random Forest (97% to 97.48%), SVM (97% to 97.39%).

• Decreased Accuracy: Decision Tree (95% to 94.29%), Naive Bayes (98% to 97.77%).

Overall, hyperparameter tuning can enhance model performance for some algo rithms while reducing it for others. The effectiveness of tuning depends on the specific algorithm and chosen hyperparameters.

Dataset	Category	Algorithm	Hyperpar-ameters used	Precision	Recall	F1-score	Accuracy
Spam_ham dataset	Spam/ ham	Logistic regression	'C': {0.1, 1, 10}, 'solver': {liblinear, lbfgs}	0.99	0.99	0.99	98.26%
				0.96	0.98	0.97	
		Decision tree	'max_depth': {None, 10, 20, 30}, 'min_samples_split': {2,10,20}	0.97	0.95	0.96	94.29%
				0.88	0.92	0.90	
		Random forest	'n_estimators': {100, 200,500}, 'max_depth': {None, 10, 20}, 'min_samples_split': {2, 10, 20}	0.98	0.99	0.98	97.48%
				0.97	0.94	0.96	
		Naïve bayes	'alpha:{0.1, 0.5, 1,0}	0.99	0.98	0.98	97.77%
				0.96	0.96	0.96	
		SVM	'C': {0.1, 1, 10}, 'gamma': {1, 0.1, 0.01}, 'kernal': {rbf, linear}	0.98	0.98	0.98	97.39%
				0.95	0.96	0.95	

Fig2:Results of Spam_Ham dataset with hyperparameters

Performance Evaluation of Classification Algorithms on the Phishing Email Dataset

In analyzing the Phishing Email dataset, five classification algorithms were evaluated: Logistic Regression, Decision Tree, Support Vector Machine (SVM), Random Forest, and Naive Bayes. Each algorithm was optimized using hyper parameter tuning. Below are the performance metrics:

1.Logistic Regression

- Accuracy: 98%(with default parameters)
- Hyperparameters: 'C': [0.1, 1, 10], 'solver': ['liblinear', 'lbfgs']
- Accuracy: 98.43% (with hyperparameters)
- Logistic Regression showed strong performance with fine-tuned parame ters.

2.Decision Tree

• Accuracy: 93.21% (with default parameters)

• Hyperparameters: ’maxdepth′ : [None,10,20,30],′ minsamplessplit′ : [2,10,20]

• Accuracy: 91.79%(with hyperparameters)

• Despite optimization, it achieved the lowest accuracy.

3.Support Vector Machine (SVM)

• Accuracy: 62% (with default parameters)

• Hyperparameters: ’C’: [0.1, 1, 10], ’gamma’: [1, 0.1, 0.01], ’kernel’: [’rbf’, ’linear’]

• Accuracy: 98.51% (with hyperparameters)

• SVM achieved the highest accuracy, showcasing its effectiveness.

4.Random Forest

• Accuracy: 97% (with default parameters)

• Hyperparameters: ’nestimators′ : [100,200,500],′ maxdepth′ : [None,10,20],′ minsamplessplit′ : [2, 10, 20]

• Accuracy: 97.17% (with hyperparameters)

• Random Forest showed robust performance.

5.Naive Bayes

• Accuracy: 96% (with default parameters)

• Hyperparameters: ’alpha’: [0.1, 0.5, 1.0]

• Accuracy: 96.37% (with hyperparameters)

• Naive Bayes performed effectively with high ccuracy.

Summary : The Support Vector Machine (SVM) achieved the highest accuracy at 98.52% , followed closely by Logistic Regression at 98.48%. Decision Tree had the lowest accuracy at 93.21%. Random Forest and Naive Bayes also demon strated strong performance with accuracies of 97.94% and 97.51%, respectively. These results emphasize the importance of model selection and hyperparameter tuning for optimal performance in phishing email detection.

Dataset	Category	Algorithm	Precision	Recall	F1 Score	Default Accuracy
Phishing email.csv	Phishing email/ safe email	Logistic regression	0.98	0.98	0.98	98%
			0.99	0.99	0.99	
		Decision tree	0.89	0.91	0.90	92%
			0.94	0.93	0.93	
		Random forest	0.99	0.94	0.96	97%
			0.96	1.00	0.98	
		Naïve bayes	0.93	0.98	0.95	96%
			0.99	0.95	0.97	
		SVM	1.00	0.02	0.03	62%
			0.62	1.00	0.76	

Fig3: Results of Phishing_Email dataset with default parameters

Dataset	Category	Algorithm	Hyperparameter used	Precision	Recall	F1_score	Accuracy
Phishing Email Dataset	Phishing /Safe	Logistic regression	'C': {0.1, 1, 10}, 'solver': {liblinear, lbfgs}	0.98	0.98	0.98	98.43%
				0.99	0.99	0.99	
		Decision tree	'max_depth': {None, 10, 20, 30}, 'min_samples_split':{2, 10,20}	0.89	0.90	0.89	91.79%
				0.94	0.93	0.93	
		Random forest	'n_estimators': {100, 200,500}, 'max_depth': {None, 10, 20}, 'min_samples_split': {2, 10, 20}	0.99	0.93	0.96	97.17%
				0.96	1.00	0.98	
		Naïve bayes	'alpha:{0.1, 0.5, 1.0}	0.93	0.98	0.95	96.37%
				0.99	0.95	0.97	
		SVM	'C': {0.1, 1, 10}, 'gamma': {1, 0.1, 0.01}, 'kernal': {rbf, linear}	0.98	0.98	0.98	98.51%
				0.99	0.99	0.99	

Fig4: Results of Phishing_Email Dataset with hyperparameter tunning

5. Future Enhancements in Email Phishing Detection

1. Advanced Feature Engineering and Selection

Improving feature engineering is essential for enhancing model performance. Future work could focus on automating feature extraction using deep learning techniques, such as Convolutional Neural Networks (CNNs) for image-based fea- tures (e.g., screenshots of emails) and Recurrent Neural Networks (RNNs) or Transformers for text-based features. Additionally, incorporating meta-features, such as user behavior patterns and historical email interaction data, could provide richer context for phishing detection.

1.Integration of Natural Language Processing

(NLP) Phishing emails often use sophisticated language to deceive recipients. Integrating advanced NLP techniques, such as BERT (Bidirectional Encoder Rep- resentations from Transformers) or GPT (Generative Pre-trained Transformer), can enhance the ability to understand the semantics and context of email con tent. These models can help in detecting subtle linguistic cues that indicate phishing attempts.

2.Real-time Detection Systems

Developing low-latency, real-time phishing detection systems is critical for timely intervention. Future enhancements could leverage edge computing and distributed systems to reduce latency and improve the scalability of detection models. Techniques such as model pruning and quantization can be applied to optimize models for real-time performance without significantly compromising accuracy.

3.User Education and Feedback Loops

Incorporating user feedback into the detection system can help refine models. Developing user-friendly interfaces that allow users to report suspected phishing emails can create a valuable feedback loop. This user-generated data can be used to continually improve the detection algorithms.

6. Conclusion

The project concluded that Support Vector Machine (SVM) and Random Forest algorithms outperformed other methods in distinguishing between phishing and legitimate emails. These algorithms exhibited superior accuracy and robustness across the evaluated datasets. Additionally,

hyperparameter tuning was found to further enhance their accuracy, leading to even better performance. Specifically, tuning the hyperparameters of SVM and Random Forest resulted in significant performance improvements. These findings suggest that utilizing these machine learning techniques, along with regular updates and fine-tuning, can greatly improve the detection of phishing emails, thereby bolstering cyber security measures. article cite.

7. References

[1] Akash S. Jayanesh B. Arul Prakash T Keerthika J., Adisvara A. et al. E mail spam detection and phishing link detection using machine learning. Department of Computer Science and Engineering, Sri Eshwar College of Engineering,Coimbatore,India, 3035, 2024.

[2] Ashraf H Aljammal, Ahmad Qawasmeh, Hani Bani Salameh, et al. Machine learning based phishing attacks detection using multiple datasets. Interna tional Journal of Interactive Mobile Technologies, 17(5), 2023.

[3] Milda Tubyte and Agne Paulauskaite-Taraseviciene. Research on phishing email detection based on url parameters using machine learning algorithms. In CEUR workshop proceedings: IVUS 2021: Information society and uni versity studies 2021: Proceedings of the 26th international conference on information society and university studies (IVUS 2021), Kaunas, Lithua nia, April 23, 2021, volume 2915, pages 18–26. CEUR-WS, 2021.

[4] Priyanka Pachpande Akshata Nevrekar Fenny Zalavadia, Shubhangi Pandey et al. Detecting phishing attacks using natural language processing and deep learning models. International Journal of Creative Research Thoughts (IJCRT), 8, 2023.

[5] Tushaar Gangavarapu, CD Jaidhar, and Bhabesh Chanduka. Applicabil ity of machine learning in spam and phishing email filtering: review and approaches. Artificial Intelligence Review, 53(7):5019–5081, 2020.

[6] A Niranjan, VK Sakhamuri, P Deepa Shenoy, and KR Venugopal. Ercrfs: Ensemble of random committee and random forest using stackingc for phish ing classification. International Journal of Emerging Trends in Engineering Research, 8(1):79–86, 2020.

[7] Sunila Godara Meenu. Phishing detection using machine learning tech niques. Int J Eng Adv Technol, 9(2), 2019