EMAIL PHISHING DETECTION USING MACHINE LEARNING

Sasirekha C[1] , Nandhini R[1], Karthiga Mai N L[2], Bhuvaneshwari R S[2], Chandra V S[2]

[1]Assistant Professor,[2]UG Scholar

Department of Electronics and Communication Engineering

K.L.N College of Engineering,

Pottapalayam, Sivagangai-630612, Tamil Nadu, India.

Abstract : ***Email phishing is a type of cyber attack that attempts to steal sensitive information by disguising as legitimate sources. Machine learning has the potential to detect email phishing attacks, and this paper presents an overview of the proposed machine learning-based approach for detection. The proposed approach involves feature extraction from emails, including message content, header information, and is used to train and test machine learning models. It uses a combination of natural language processing and supervised learning algorithms to classify incoming emails as either legitimate or phishing attempts. The results show that the proposed approach achieves high accuracy and outperforms existing approaches, and can be used by organizations and individuals to improve their email security.***

Keywords: Phishing email, Random Forest Classifier

1.INTRODUCTION

Phishing attacks are a type of cyber attack that target email users. They are designed to trick users into providing sensitive information such as passwords, credit card details, and other personal information by disguising the email as a legitimate entity. Email is a digital communication method used to exchange messages over the internet.

Phishing attacks use email as the primary mode of communication, impersonating a trusted source. The email contains a message that urges the recipient to take immediate action, such as clicking on a link, downloading an attachment, or entering their personal information. It is one of the most common forms of cyber attacks, and its success relies on the victim's ability to distinguish between legitimate and fake emails.

Machine learning techniques are used to detect email phishing attempts by training a model with a labeled dataset of phishing and legitimate emails. The model is then used to classify new emails as either phishing or legitimate based on their features and characteristics.

Phishing attacks can take various forms, including:

Spear Phishing: Attackers use publicly available information and social engineering tactics to create personalized messages that appear to be legitimate, making them appear more convincing than they are.

Whaling: Attackers use spear phishing tactics to target high-level executives or decision-makers in organizations, targeting individuals with more authority and access to sensitive information.

Clone Phishing: Attackers create an email that appears to be from a trusted source and replace the attachment or link with a malicious one.

Pharming: Attackers redirect users from legitimate websites to fake ones, asking them to enter sensitive information that they will capture**.**

Phishing attacks are a significant threat to online security and can result in identity theft, financial loss, and damage to an individual's or organization's reputation. To protect oneself, it is important to verify the sender's identity, avoid clicking on suspicious links or attachments, and report any suspicious activity to the relevant authorities. The email phishing detection using machine learning project can help organizations and individuals to better protect themselves against phishing attacks and safeguard their sensitive information. By automating the process of detecting phishing emails, the system can save time and resources and provide more effective protection against cyber threats. This project highlights the importance of using advanced technologies to combat cyber threats and ensure online security.

2. EXISTING SYSTEM

2.1 LONG SHORT-TERM MEMORY

LSTM (Long Short-Term Memory) is a type of recurrent neural network architecture used in machine learning for sequence modeling and prediction tasks. It is composed of cells that contain three "gates" - an input gate, an output gate, and a

forget gate - that regulate the flow of information into and out of the cell. The input gate determines which information to update in the cell, the output gate determines which information to output, and the forget gate determines which information to discard. The gates are controlled by activation functions that learn to open or close them based on the input data.
LSTM models can be used to detect email phishing by analyzing text content and identifying patterns indicative of attacks.

The advantages of LSTM are,

- It can be trained to detect phishing attacks using a variety of features, such as email headers, content, and metadata.
- LSTM can identify patterns in phishing emails that can be used to deceive users.

And the disadvantages of LSTM are:

- LSTM models require large amounts of training data to achieve good performance, making them expensive to train.
- LSTM models may struggle to generalize to new data that is significantly different from training data.
- LSTM models may require additional techniques to handle imbalanced datasets.
- LSTM models are more complex and difficult to interpret.

3. PROPOSED SYSTEM

3.1 RANDOM FOREST CLASSIFIER

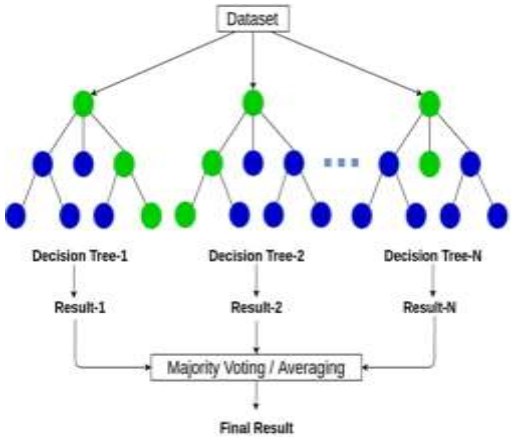
Data science provides various classification algorithms, such as Support vector machine, Naive Bayes, Logistic regression, and Decision tree, but Random forest is the top of the classifier hierarchy.

Random Forest is an ensemble learning method that uses multiple decision trees and outputting class mode (classification) or mean prediction. It uses a random subset of training data and features to reduce overfitting and improve performance.
The key steps involved in building a Random Forest are as follows:

- Randomly select a subset of the training data.
- Randomly select a subset of features for each tree.
- Build a decision tree using the selected data and features.
- Repeat steps 1-3 to create a forest of decision trees.

Predicting a new instance requires passing it through each tree in the forest and taking the mode of predicted classes or mean prediction.

Random Forest is a popular machine learning algorithm used in email phishing detection due to its high accuracy, robustness, and ability to handle large datasets with many features.



Flow diagram for Random Forest Classifier

Advantages of Random Forest classifier are:

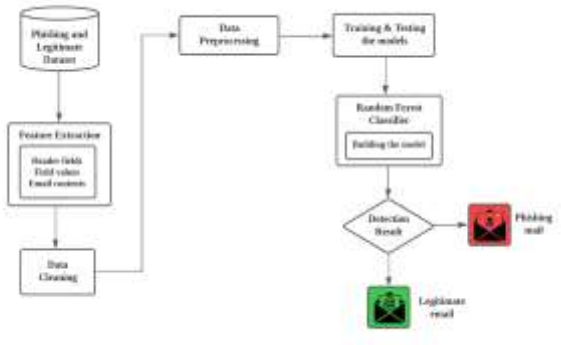
- Random Forest classifier can be trained faster and more efficiently, making it a good choice for applications with speed.
- Random Forest classifiers can better generalize to new or unseen data than LSTM models.
- Random Forest classifier can handle imbalanced datasets well.
- Random Forest classifier provides feature importance scores to help identify which features are most relevant for phishing detection.
- Random Forest is a powerful tool for email phishing detection, as it can handle large datasets and features.

3.2 IMPLEMENTATION

Random Forest classifier is used to detect email phishing by creating a large number of decision trees and combining their predictions to generate a final output.

Random Forest is an algorithm that uses a large number of decision trees to classify emails based on a random subset of the available data and features. During training, the algorithm builds multiple decision trees, each of which learns to classify emails based on a different subset of the available features. The predictions of these individual trees are then combined using a voting mechanism, where the final

classification is determined by the majority vote of the trees. To use Random Forest for email phishing detection, the first step is to collect a dataset of emails that have been labeled as either legitimate or phishing..



Block diagram for Email Phishing Detection using Machine Learning

3.3 METHODOLOGY

Here is a methodology that can be used for implementing a random forest classifier for email phishing detection:

3.3.1 DATA PREPARATION, CLEANING AND PREPROCESSING

DATA PREPARATION: The first step in data preparation is to gather the data needed to train the phishing detection model. This data could include phishing emails, legitimate emails, or a combination of both. The data should be representative of the types of emails the detection model is likely to encounter in the real world.

Data labeling: The labeling process is important because it provides the ground truth for the detection model to learn how to distinguish between phishing and legitimate emails.

Data splitting: The dataset needs to be divided into training, validation, and testing sets to train the detection model, tune its hyper parameters, and evaluate its performance on unseen data.

Feature extraction: Features such as the sender's address, subject, body, and attachments must be extracted from emails in the dataset to be used in the detection model.

DATA CLEANING: Data cleaning is the process of correcting any errors or inconsistencies in a dataset to improve its performance in email phishing detection.

Removing duplicates: Duplicate emails should be removed from the dataset to avoid biasing the detection model's training.

Removing irrelevant data: Data that is not useful for the detection model should be removed and standardized to ensure consistency across the dataset.

The most important details in this text are that email addresses should be converted to lowercase, and that missing data should be filled in using imputation techniques to avoid biasing the detection model's training. By properly preparing and cleaning the data, the phishing detection model can be trained on a high-quality dataset, leading to better accuracy and more effective detection.

DATA PREPROCESSING: The most important details in this text are that email messages must be preprocessed to prepare them for analysis, such as removing email headers, extracting the email body, and converting the text to a format that can be used by machine learning algorithms. Finally, the labeled email messages can be used to train and evaluate machine learning models that can detect phishing attempts with high accuracy. This involves using algorithms such as decision trees, random forests, or neural networks to classify new email messages as either phishing or legitimate.

4. RESULT AND DISCUSSION

Random forest classifier is a machine learning algorithm used to classify applications as either phishing or non-phishing.

Random forest is known for its high accuracy and is used for phishing detection, as it works well with large datasets and can handle non linear relationships between features

High accuracy: Random Forest models are known for their high accuracy and ability to handle noisy data.

Fast training and prediction: Random Forest models can be trained quickly and make predictions quickly, making them useful for real-time applications such as email phishing detection.

Robustness: Random Forest models are less prone to overfitting and can handle missing or corrupted data, making them a good choice for real-world applications.

Feature importance: Random Forest models can provide information on the importance of each feature

in the prediction, which can help understand the underlying factors contributing to phishing emails.

Random Forest is a powerful and flexible machine learning algorithm that can be used to detect email phishing.

5. CONCLUSION

Random forest is a machine learning algorithm that can analyze a large number of variables and create decision trees to identify patterns and classify emails as legitimate or phishing attempts. It can reduce the risk of falling victim to phishing scams by flagging suspicious emails and preventing them from reaching users' inboxes. Future work can include deep learning to enhance the effectiveness, robustness and accuracy of the algorithm.

6. REFERENCES

[1] K. Zetter, L. Matsakis, I. Lapowsky, G. Graff, E. Dreyfuss, and L. Newman, "Researchers uncover RSA phishing attack, hiding in plain sight," WIRED, 2018.[Online].Available:https://www.wired.com/2011/08/how-rsa-got-hacked

[2] L. Matsakis, I. Lapowsky, G. Graff, E. Dreyfuss, and L. Newman, "Why the DNC thought a phishing test was a real attack," WIRED,2018.[Online].Available:https://www.wired.com/story/dnc-phishingtest-votebuilder

[3] M. Alsharnouby, F. Alaca, and S. Chiasson, "Why phishing still works: User strategies for combating phishing attacks," Int. J.Hum.-Comput. Stud., vol. 82, pp.69–82,2015.[Online]. Available:10.1016/j.ijhcs.2015.05.005.

[4] T. Jagatic, N. Johnson, M. Jakobsson, and F. Menczer, "Social phishing," Commun. ACM, vol. 50, no.10, pp. 94–100, 2007. [Online].Available: 10.1145/1290958.1290968.

[5] N. Arachchilage, S. Love, and K. Beznosov, "Phishing threat avoidance behaviour: An empirical investigation," Comput. Hum. Behav.vol. 60, pp. 185–197, 2016.

[6] T. Nikolaos, V. Nikos and M. Alexios, "Browser blacklists: The utopia of phishing protection," in Proc. E-Business Telecommun, 2014, pp. 278–293.

[7] W. Khan, M. Khan, F. Bin Muhaya, M. Aalsalem, and H. Chao, "A comprehensive study of email spam botnet detection," IEEE Commun. Surv. Tuts., vol. 17, no. 4, pp. 2271–2295, 2015.

[8] S. Jeeva and E. Rajsingh, "Phishing URL detection-based feature selection to classifiers," Int. J. Electron. Secur. Dig. Forensics, vol. 9, no. 2, 2017, Art. no. 116.

[9] J. Chaudhry and R. Rittenhouse, "Phishing: Classification and countermeasures," in Proc. Int. Conf. Multimedia, 2016, pp. 28–31.

[10] H. Che, Q. Liu, and L. Zou, "A content-based phishing email detection method," in Proc. IEEE Int. Conf. Softw. Quality Rel. Secur. Companion, 2017, pp. 415–422.

[11]C. Tan, K. Chiew, K. Wong, and S. Sze, "PhishWHO: Phishing webpage detection via identity keywords extraction and target domain name finder," Decis. Support Syst., vol. 88, pp. 18–27, 2016.