SOCIAL ENGINEERING

|  |  |  |  |
| --- | --- | --- | --- |
| Kartik Kumar  B.E CSE IS *Chandigarh University* Gharuan, Punjab, India [21BCS8403@cuchd.in](mailto:20BCS3717@cuchd.in) | Pratik Mukherjee  B.E CSE IS *Chandigarh University* Gharuan, Punjab, India [21BCS8404@cuchd.in](mailto:21BCS8403@cuchd.in) | Krishma  B.E CSE IS *Chandigarh University* Gharuan, Punjab, India [21BCS8405@cuchd.in](mailto:21BCS8405@cuchd.in) | Priyanka Jammwal  B.E CSE IS *Chandigarh University* Gharuan, Punjab, India priyanka.e15553@cumail.in |

**Abstract- *Social engineering has emerged as a significant threat vector in the realm of cybersecurity, exploiting human psychology to manipulate individuals into divulging sensitive information or performing actions that compromise security. This research delves into one of the most prevalent and insidious forms of social engineering: phishing. The project abstract provides an overview of the study's objectives, methodologies, and anticipated contributions. Phishing is a deceptive practice wherein cyber attackers masquerade as legitimate entities through emails, messages, or websites, aiming to trick recipients into revealing confidential information such as passwords, financial details, or personal data. The goal of this research is to comprehensively analyze the concept of phishing within the context of social engineering, shedding light on its underlying mechanisms, tactics, and potential countermeasures. The study employs a multi-faceted approach, combining literature review, case studies, and empirical analysis. By examining real-world phishing attacks and their outcomes, the project seeks to uncover the psychological triggers that make individuals susceptible to such manipulation. Additionally, technical aspects of phishing campaigns, such as email spoofing and website replication, will be explored to understand the intricacies of attack execution.***

***Keywords—Phishing; URL features; Social Engineering, Phishing Detection, Confidential Information.***

# INTRODUCTION

The evolution of communication technologies and the advent of digitalization

–has enabled life to become faster and more accessible, especially during the lockdown resulting from the COVID-19 pandemic when all transactions and life necessities were procured through the internet, i.e., shopping and transactions, rather than having to do so in person. To fulfill your daily needs, you can simply open your smart device, and search for the website you want. For example, a pharmacy, shopping store, learning platform, or bookstore. Additionally, E- services are on the rise, which increases the chances of attackers gaining or misusing a user's personal information, such as their name, phone number, identification, and credit card number. This has resulted in a range of online threats and cyberattacks being faced by users every day as a consequence. One of them is phishing which is done mainly through emails. Phishing is a cybercrime that employs social engineering and technical subterfuge to steal victims' personal identity data or financial account credentials.

Phishing is a type of online fraud where attackers create fake versions of reputable websites to lure individuals into divulging confidential information such as usernames, passwords, banking or credit card details, and other sensitive credentials. These phishing URLs may be sent to consumers through email, instant messages, or text messages.

The Phishing detection system is designed to deal with cyberattacks caused due to fake web links that steal confidential information using malware viruses, etc. This first of all dataset of phishing and legitimate emails will be collected and processed further for analysis. A particular machine learning algorithm is used either a decision tree or a neural network to analyze the dataset. ML is an arm of artificial intelligence and it is analogous to (and frequently overlaps with) computational measurements, that also concentrate on making predictions with the use of PCs. Machine learning has a solid relationship with scientific improvement, which tells methods, hypotheses, and utilization regions of the field. ML is sometimes, in a while combined with data mining, but the data mining subfield focuses more on preparatory information investigation and is called unsupervised learning.ML can likewise be unsupervised and utilized to learn and set up patterns for various entities and then used tofind important anomalies.

The applications of machine learning (ML) methods in cybersecurity are rising more than ever before. Beginning from IP traffic categorization, and separating malicious traffic for intrusion detection, Machine learning is one of the best answers that can impact zero-day attacks. New exploration is being done through the utilization of measurable traffic characteristics and ML techniques. The word phishing was introduced in the year 1987. Phishing is an online thievery that robs an individual’s private data and identity data. It is a sort of extortion where the assailant gets complete access to other individuals’ private data. As the number of phishing attacks has increased, numerous solutions have been proposed to address this issue. As far as building a framework that guarantees a solution to a phishing attack is concerned, several things can be done. It is also possible to use various

other methods for detecting phishing attacks. These include black lists, fuzzy rule-based detection, white list-based detection, canteen- based detection, machine learning-based detection, heuristic detection, and image- based detection methods. A variety of other studies discuss different methods and techniques for detecting different types of phishing attacks. Phishing sites look like genuine websites, and many people have trouble recognizing such websites. Some antiphishing techniques are built into some of the browsers.

# BACKGROUND

* 1. Phishing Detection

A URL-based phishing attack is carried out by sending malicious links, that seem legitimate to the users, and tricking them into clicking on it. In phishing detection, an incoming URL is identified as phishing or not by analyzing the different features of the URL and is classified accordingly. To determine whether a URL is genuine or a phishing attempt, different machine learning algorithms are trained on distinct datasets of URL characteristics.

* 1. Phishing detection Approaches

As part of the List Based Approach, there are two lists, whitelist and blacklist, in order to classify legitimate and phishing URLs. Website access is granted only if the URL is whitelisted. In blacklist is used. The Heuristic-Based approach involves analyzing the structure of a phishing URL. A pattern of phishing URLs is created. URLs are classified according to this pattern. By comparing the visual similarity between pages of websites, the visual similarity of the URL plays a significant role in classifying websites accurately. To determine whether websites are fraudulent or not, a server-side

analysis is conducted. This data is then compared to the original website using image processing techniques, as fake web pages are often designed to closely resemble the real ones. Image processing techniques are more effective in identifying minor discrepancies that may be difficult for users to notice.

The content-Based approach analyses pages' content. Features are extracted from both page content and third-party services, such as search engines and DNS servers, using this method. These features may include words like brand names that attackers incorporate into the URL to create the illusion of a genuine website. The presence of these words at different positions in the URL is assigned weights to determine their significance. The most probable words are chosen and then sent to Yahoo search to return the domain name with the highest frequency among the top 30 results. The domain name owners are compared to decide if the website is phishing or not. In this, a logo image was utilized to authenticate the identity of web pages by comparing genuine and fraudulent web pages.

By using a fuzzy rule-based approach, we can process ambiguous variables and add human experts to classify those variables and their relationships. A set of predefined rules and metrics are employed to classify web pages based on the degree of phishing present within them using this approach.From the experimental results in the paper, fuzzy logic systems have lower features. If a fuzzy logic algorithm is affected by irrelevant features, the classifier's effectiveness will decrease and vice-versa.

The Machine Learning-based approach involves the creation of machine learning models that utilize supervised learning algorithms to classify a given URL as either legitimate or a phishing attempt. Different algorithms are trained on a dataset and then tested to learn each model's performance. The

performance of the model is directly impacted by any inconsistencies in the training data. Nonetheless, this approach offers effective phishing detection techniques and is a significant area of research. Numerous papers have been published on the subject of machine learning-based phishing detection.

* 1. Machine learning Algorithm
* Decision Tree Algorithm

Machine learning technology's most widely used algorithm. Decision tree algorithm is easy to understand and implement. Decision tree begins its work by choosing the best splitter from the available attributes for classification which is considered as a root of the tree. The algorithm continues to build tree until it finds the leaf node. The decision tree creates training model which predicts target value or class in tree representation. Each internal node of the tree belongs to the attribute and each leaf node belongs to class label. In decision tree algorithms, nodes are calculated using the Gini index and the information gain method.

* Random Forest Algorithm

Random forest algorithm is one of the most powerful algorithms in machine learning technology and it is based on concept of decision tree algorithm.

A random forest algorithm creates a forest with a number of decision trees. Detection accuracy increases with a high number of trees. The bootstrap method is used to create trees. With bootstrap method, features and samples of dataset are randomly selected and replaced with new ones to construct a single tree.

As with decision tree algorithm, random forest algorithm will select the best splitter from randomly selected features. Random forest algorithm also uses Gini index and information gain methods to choose the best splitter for classification. This process will continue until random forest creates n number of trees. Each tree in forest predicts the target value and then algorithm calculates the votes for each predicted target. As a final prediction, random forest algorithm considers the most voted predicted target.

* Neural Network

A neural network is a type of machine learning model that is inspired by the structure and function of the human brain. It consists of a large number of simple processing units called neurons, which are organized into layers. The neurons in each layer are connected to neurons in the previous and subsequent layers, forming a network. Each neuron receives input from other neurons, performs a simple computation, and passes its output to other neurons in the next layer. The structure of a neural network is typically organized into three main types of layers: input layer, hidden layer(s), and output layer. The input layer takes in the input data, and the hidden layer(s) perform a series of computations on the input data to extract relevant features. The output layer then produces the final output, which can be a classification label, a prediction, or a probability distribution. The computations performed by the neurons are based on weights and biases. The weights determine the strength of the connection between neurons, while the biases determine the threshold at which a neuron is activated. During training, the weights and biases of the neurons are adjusted to minimize the difference between the predicted output and the actual output. This

is typically done using an optimization algorithm such as stochastic gradient descent. One of the key advantages of neural networks is their ability to learn complex patterns in data. This is achieved by adjusting the weights and biases during training to produce more accurate predictions. The process of adjusting the weights and biases is known as backpropagation, and it involves propagating the error from the output layer back through the network to adjust the weights and biases in the hidden layers. Neural networks have become increasingly popular in recent years due to their ability to make accurate predictions on a wide range of tasks, including image classification, natural language processing, and speech recognition. For example, a neural network can be trained to recognize images of cats and dogs by learning the features that distinguish them, such as the shape of their ears, the color of their fur, and the texture of their skin.

# LITERATURE SURVEY

A literature survey is an insightful article that presents the existing information including considerable discoveries just as theoretical and methodological commitments to a specific topic. With the advent of email, it has become easier for people to communicate, leading to unsolicited bulk emails, especially phishing attacks via emails. Various anti- phishing techniques have been developed to solve the problem of phishing attacks. This paper focuses on determining whether important emails are spam or if they are important emails One of the primary considerations for classification is the representation of messages. It is crucial to determine which features should be utilized and how they should be employed in the categorization process. It is well known that many researchers have used artificial

intelligence to build intelligent systems, and many of them have applied deep learning to cybersecurity systems as well. Using an optimal feature selection technique and also using a neural network to detect phishing websites (OFS-NN), a very effective model of phishing website detection is presented. In this proposed model, an index called feature validity value(FVV) has been generated to check the effects of all those features on the detection of such websites.

Now, based on this newly generated index, an algorithm is developed to find from the phishing websites, the optimal features. This selected algorithm will be able to overcome the problem of over-fitting of the neural network to a great extent. These optimal features are then used to build an optimal classifier that detects phishing URLs by training the neural network.

Feature engineering plays a vital role in finding solutions for the detection of phishing websites, although the accuracy of the model greatly will be based on knowledge of the features. though the features taken from all these various dimensions are understandable, the limitation lies in the time taken to collect these features. To fix this drawback, the authors have proposed a multidimensional phishing detection feature approach that concentrates on a rapid detection technique by making use of deep learning (MFPD).A three-phase detection method, known as Web Crawler based Phishing Attack Detector (WC-PAD), has been proposed to detect phishing attempts with precision by leveraging deep learning techniques, referred to as MFPD.This takes the web’s content, traffic, and URL as input features. Now considering these features, classification is done.

PhishingNet is an approach based on deep learning for detecting phishing URLs quickly. A detection system was developed to match the dynamic environment and phishing websites. In this approach, different types of

distinctive features are taken into account from the source code of webpages and URLs, so there is no need for the involvement of a third party.A method called parse tree validation has been proposed to find if a webpage is phishing or legitimate. This is an innovative approach to finding such websites by intercepting every hyperlink of a present page through the API of Google and developing a parse tree from all those hyperlinks that were intercepted. This method commences parsing from the root node and employs the Depth-First Search (DFS) algorithm to examine whether any child node contains the same value as the root node.

For the detection of phishing websites by URL method, the Random Forest classifier was used as a solution. The tool combines the collection, validation, and detection of phishing websites. This online tool monitor in real-time the blacklist of PhishTank, and validates and detects phishing websites. This framework is known as "Fresh-Phish", which generates machine learning data from phishing websites. By using 30 various features of the website which can be queried using Python, a very large dataset is built and the various ML classifiers are analyzed against this generated dataset to find out which has the highest accuracy. This model analyzes both accuracy and the time taken by the model to train.

Qabajeh et al. (Qabajeh et al., 2018)-A recent study compared conventional and automated phishing detection techniques. Conventional approaches to anti-phishing include raising awareness, user education, periodic training or workshops, and legal measures. Automated anti-phishing techniques include list-based and machine learning-based approaches. The paper examines the similarities, advantages, and disadvantages of these approaches from the user and performance perspectives.

The study concluded that machine learning and rule induction are effective in combating phishing attacks.. The study's limitations are that it is based on a review of only 67 research items, and it does not include Deep Learning techniques for the detection of phishing websites in the study.

Kunju et al. (2019) employed a survey method to identify phishing attacks and proposed several solutions and methodologies for detecting such attacks. According to the research, many of the proposed solutions were insufficient in solving phishing attacks. The literature in this work includes only 14 studies between 2007 and 2019. The study discusses only Machine Learning techniques for phishing website detection.

Athulya & Praveen (Athulya and Praveen, 2020) addressed different phishing attacks, phishers' latest phishing tactics, and anti- phishing strategies.Furthermore, the article strives to increase awareness of phishing attacks and highlight strategies employed for detecting phishing attempts. According to this study, the effective way to prevent phishing attacks is to educate users about the different types of phishing attacks. Users can choose the right security software tools or applications to detect phishing attacks, such as anti-phishing browser extensions. The literature in this work consists of nine research items. The study does not include Deep Learning techniques for phishing website detection.

Arshad et al. (Arshad et al., 2021) presented different types of phishing and anti-phishing techniques in their study. The SLR evaluated that phone phishing, Email Spoofing, spear phishing, and Email Manipulation are the most commonly used phishing techniques.

According to this study, the highest accuracy was achieved through Machine Learning approaches. The research is limited to 20 studies.

Catal et al. (Catal et al., 2022) performed a systematic literature review, answering nine research questions. The study's main aim is to identify, assess, and synthesize Deep Learning approaches for phishing detection. According to this study, supervised machine learning algorithms were applied to 42 studies out of 43. The most used algorithm was DNN, and the highest performance was given by DNN and Hybrid DL algorithms.

# METHODOLOGY

*Data collection*

Collecting a dataset of legitimate and phishing emails. Data collection refers to the process of gathering information from various sources, such as websites, databases, or sensors, for a specific purpose or research project. In the context of phishing detection using machine learning, data collection involves collecting a representative dataset of both legitimate and phishing emails that can be used to train and test a machine learning model. The data can be obtained from various sources, such as publicly available datasets, online repositories, or through manual collection by researchers. The collected data should be diverse and balanced in terms of the number of legitimate and phishing emails, and should represent different types of phishing attacks and techniques. The quality and quantity of the collected data is crucial for the success of the machine learning model and its ability to accurately detect phishing emails in real- world scenarios.

*Data pre-processing*

Cleaning and pre-processing the dataset by removing irrelevant information, such as email headers or signatures, and converting the remaining text into numerical features. Data processing refers to the transformation of raw data into a more meaningful form that can be used for analysis or decision-making. In the context of phishing detection using machine learning, data processing involves various steps such as cleaning, pre- processing, feature engineering, and selection.

Cleaning refers to the removal of any noise or irrelevant data from the dataset, such as duplicates, irrelevant features, or incomplete records. Pre-processing involves transforming the data into a suitable format for machine learning algorithms, such as converting categorical variables into numerical values, scaling the data, and handling missing values.

Feature engineering involves selecting relevant features or variables from the dataset that can help the machine learning algorithm to differentiate between legitimate and phishing emails. This may include features such as the sender's email address, subject line, content, or metadata. Feature selection involves reducing the number of features to only those that are most important or informative for the model.

Data processing is an essential step in the development of a machine learning-based phishing detection system as it helps to improve the quality of the dataset and the accuracy of the model's predictions.

*Feature extraction*

Identifying and selecting relevant features that can discriminate between legitimate and phishing emails. Feature selection is the process of selecting a subset of relevant and informative features (also known as variables

or attributes) from a larger set of features in a dataset. In machine learning, the choice of features used to train a model is critical to the model's performance. Irrelevant or redundant features can lead to overfitting or poor generalization, while informative features can improve the model's accuracy and interpretability. Feature selection techniques aim to identify and retain only the most relevant and informative features for a given task, while discarding the rest.

This helps to reduce the complexity of the model and improve its efficiency, as well as improve the quality of the predictions made by the model.

The below-mentioned category of features are extracted from the URL data:

* 1. *Address Bar based Features* In this category 9 feature are extracted.
  2. *Domain based Features*

In this category 4 feature are extracted.

* 1. *HTML & Javascript based Features* In this category 4 feature are extracted

*Training*

Using machine learning algorithms, such as decision trees, logistic regression, or neural networks, to train a classification model on the selected features. In machine learning, training is the process of teaching a model to make accurate predictions by exposing it to labeled examples (data with known inputs and outputs). The goal of training is to learn a set of parameters or weights that allow the model to map inputs to outputs accurately. During training, the model iteratively adjusts its parameters based on the error between its predicted outputs and the actual outputs in the training data. This process is known as optimization or parameter estimation.

The model learns from its mistakes and updates its parameters to reduce prediction errors. Before starting the ML model training, the data is split into 80-20 i.e., 8000 training samples & 2000 testing samples.

This data set comes under a classification problem, as the input URL is classified as phishing (1) or legitimate (0). The supervised machine learning models (classification) considered to train the dataset in this project are:

* Decision Tree
* Random Forest
* Multilayer Perceptrons
* XGBoost
* Autoencoder Neural Network
* Support Vector Machines

The quality of the training data, the choice of model architecture, and the training algorithm used are all critical factors that can affect the performance of the trained model. The trained model can then be used to make predictions on new, unseen data.

*Evaluation*

Testing the trained model on a separate validation dataset and evaluating its performance using metrics such as accuracy, precision, recall, and F1-score. The evaluation of a phishing detection system involves assessing the performance of the system in detecting and correctly classifying phishing attacks. This is typically done by using a set of evaluation metrics such as accuracy, precision, recall, F1 score, and area under the ROC curve (AUC).

To evaluate a phishing detection system, a dataset of known phishing and legitimate emails or URLs is usually used. The system is trained on a subset of the dataset, and the remaining data is used for testing. The

From the dataset, it is clear that this is a supervised machine-learning task.

These are two major types of supervised machine learning problems, called classification and regression.

system's performance is then evaluated by comparing its predictions against the known classifications in the test dataset.

The evaluation of a phishing detection system is important because it allows researchers and practitioners to determine the effectiveness of the system and identify any areas for improvement.

*Deployment*

Integrating the trained model into a real-time phishing detection system that can classify incoming emails as either legitimate or phishing. Deployment of a phishing detection system using machine learning involves making the system available for use in a production environment. This typically involves integrating the system with existing software and hardware infrastructure, and ensuring that the system is scalable, reliable, and secure.

The deployment process may involve additional testing and validation to ensure that the system works as intended in real- world scenarios. The system may also need to be continuously monitored and updated to maintain its effectiveness in detecting new and evolving phishing attacks.

Deployment also involves considerations such as user interfaces, documentation, and training for end-users and administrators. It is important to ensure that the system is easy to use and understand, and that users have the necessary knowledge and skills to effectively utilize the system

.

Fig.3 Correlation heatmap

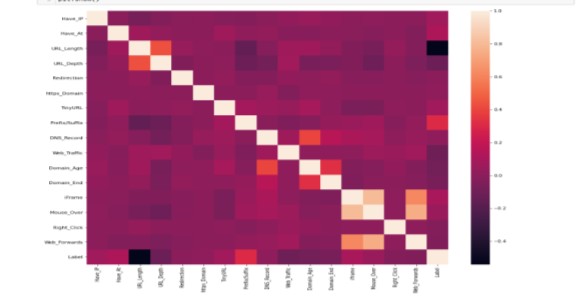
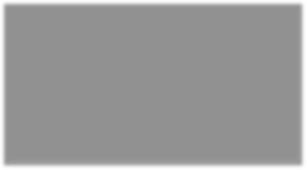
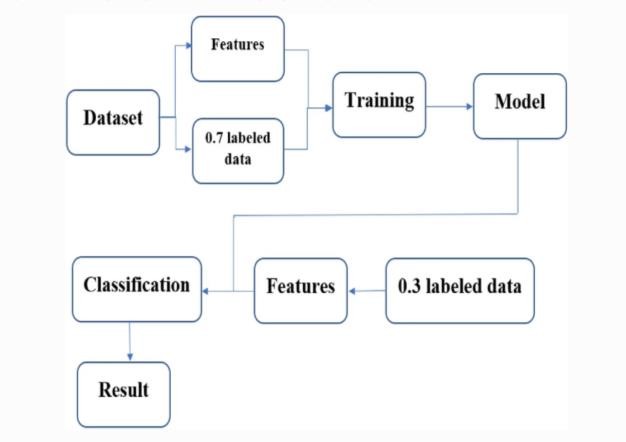
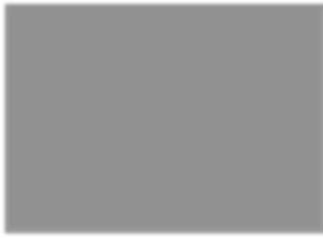
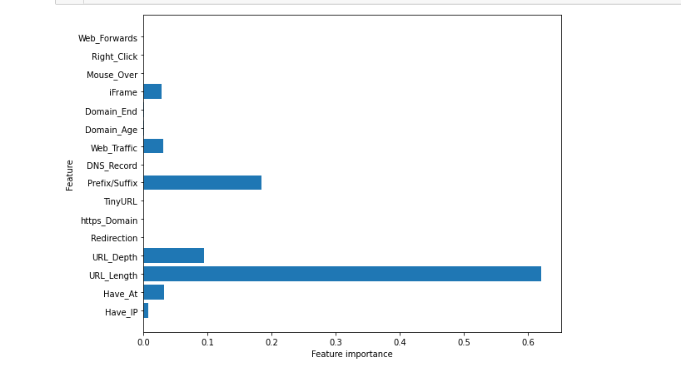
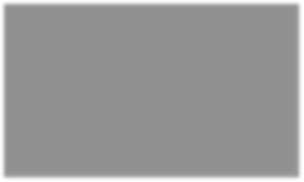
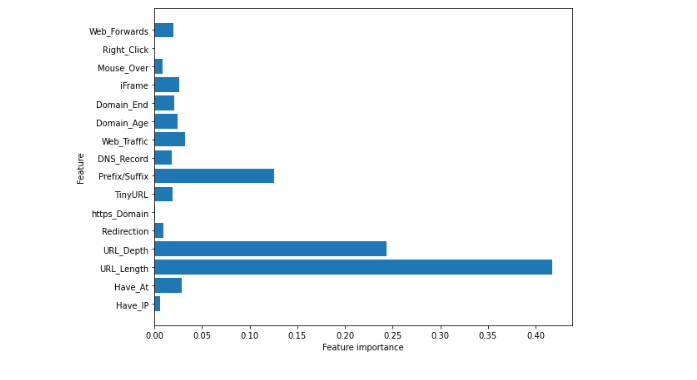
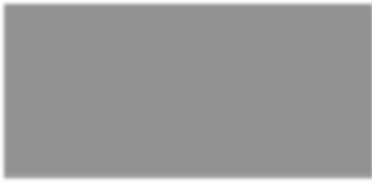


Fig 1- Flowchart of the phishing detection system.



# Implementation and Result



Five data files are used for extraction of data and to analyse which url is phishing and which is legitimate.

* 1. Benign\_list\_big\_final
  2. online-valid
  3. legitimate
  4. urldata
  5. phishing

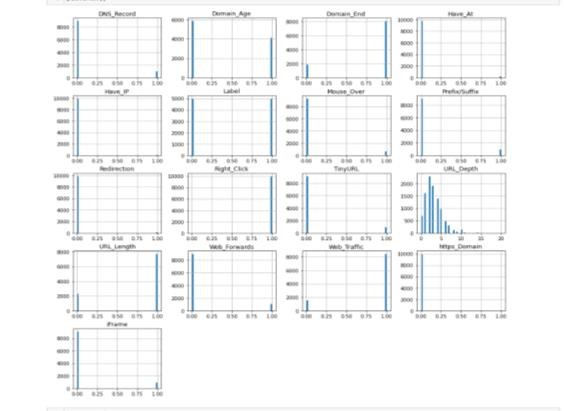
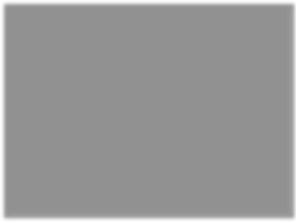


Fig. 2 Plotting the data distribution

Fig 4. Checking the feature importance in the Model using decision tree classifier

Fig 5. Checking the feature importance in the Model using Random Forest classifier.

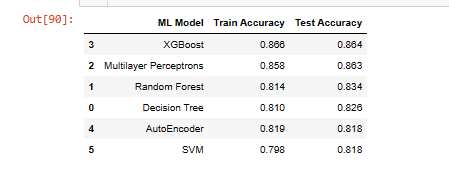
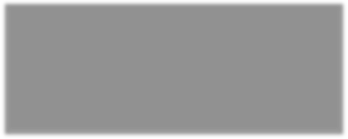
Fig. 6 Comparison of models

# CONCLUSION

In recent years, phishing emails have become a prevalent issue. Phishing email attacks are intelligently crafted social engineering email attacks in which victims are conned by email to provide sensitive information and then directly send it to the phisher. It is more likely that young users will fall prey to phishing attacks because of their naiveté. Additionally, users with the trait of agreeable behaviour are more likely to be lured by phishing scams than those with the trait of disagreeable behaviour. Women are more likely to provide their personal and financial details to fraud emails and websites. There is a causal relationship between gender and social engineering that can be explained by internet usage patterns. So, detection of that type of email is necessary. There are numerous techniques for detecting phishing emails. However, there are a few limitations, such as a low level of accuracy. The content material may be the same as legitimate email so it cannot be detected, the detection rate is not high. This study utilized machine learning techniques to attain superior outcomes and capture inherent attributes of the email text and other characteristics for classifying emails as either phishing or non-phishing. The results of this research have improved the

accuracy of detecting phishing emails. A comparison between three supervised datasets was conducted between these classifiers. Numerous machine learning algorithms exhibit efficient classification with strong performance metrics. In this article, we conducted a study of the phishing detection process and phishing detection techniques presented in recent research literature. This study can serve as a reference for new researchers to comprehend the process and formulate more precise phishing detection systems.

# REFERENCES



* 1. What is Phishing? : [https://www.phishing.org/what-is-](https://www.phishing.org/what-is-phishing) [phishing](https://www.phishing.org/what-is-phishing).
  2. Sarker. I.H deep Cybersecurity: A Comprehensive Overview from neural network and deep learning perspective. SN *Comput. Sci.*2021,2,154.
  3. Kim.D.,Kim, Y.-H., Shin,D.: Fast attack detection system using log analysis and attack tree generation clust. Comput.22(1), 1827-1835(2019).
  4. Phishing Detection using Machine Learning-based URL Analysis: A Survey-

[https://www.ijert.org/research/phishing](https://www.ijert.org/research/phishing-detection-using-machine-learning-based-url-analysis-a-survey-IJERTCONV9IS13033.pdf)

[-detection-using-machine-learning-](https://www.ijert.org/research/phishing-detection-using-machine-learning-based-url-analysis-a-survey-IJERTCONV9IS13033.pdf) [based-url-analysis-a-survey-](https://www.ijert.org/research/phishing-detection-using-machine-learning-based-url-analysis-a-survey-IJERTCONV9IS13033.pdf) [IJERTCONV9IS13033.pdf](https://www.ijert.org/research/phishing-detection-using-machine-learning-based-url-analysis-a-survey-IJERTCONV9IS13033.pdf).

* 1. An intelligent cyber security phishing detection system using deep learning techniques.
  2. [https://link.springer.com/article/1](https://link.springer.com/article/10.1007/s10586-022-03604-4) [0.1007/s10586-022-03604-4](https://link.springer.com/article/10.1007/s10586-022-03604-4).
  3. [https://www.academia.edu/4323094](https://www.academia.edu/43230944/A_REPORT_on_DETECTION_OF_PHISHING_WEBSITE_USING_MACHINE_LEARNING) [4/A\_REPORT\_on\_DETECTION\_OF\_PH](https://www.academia.edu/43230944/A_REPORT_on_DETECTION_OF_PHISHING_WEBSITE_USING_MACHINE_LEARNING) [ISHING\_WEBSITE\_USING\_MACHINE](https://www.academia.edu/43230944/A_REPORT_on_DETECTION_OF_PHISHING_WEBSITE_USING_MACHINE_LEARNING)

[\_LEARNING.](https://www.academia.edu/43230944/A_REPORT_on_DETECTION_OF_PHISHING_WEBSITE_USING_MACHINE_LEARNING)

* 1. Panos Louridas and Christof Ebert. Machine learning. IEEE Software, 33:110– 115, 09 2016.

9.

https:/[/www.ncbi.nlm.nih.gov/pm](http://www.ncbi.nlm.nih.gov/pmc/article)c[/article](http://www.ncbi.nlm.nih.gov/pmc/article) s/PMC8504731/

1. Ariyadasa, S.; Fernando, S.; Fernando, S. Detecting phishing attacks using a combined model of LSTM and CNN. *Int. J. Adv. Appl. Sci.* 2020, *7*, 56– 67.
2. https:/[/www.kore](http://www.koreascience.or.kr/artic)a[science.or.kr/artic](http://www.koreascience.or.kr/artic) le/JAKO201905959996575.page
3. Tang, L.; Mahmoud, Q.H. A deep learning-based framework for phishing website detection. *IEEE Access* 2022, *10*, 1509–1521.
4. URL 2016|Datasets|Research|Canadian Institute for Cybersecurity |UNB. Unb.ca. 2022. Available online:

https:/[/www.unb.c](http://www.unb.ca/cic/datasets/url-)a[/cic/datasets/url-](http://www.unb.ca/cic/datasets/url-) 2016.html

1. Do, N.Q.; Selamat, A.; Krejcar, O.; Herrera-Viedma, E.; Fujita, H. Deep Learning for Phishing Detection: Taxonomy, current challenges and Future Directions. *IEEE Access* 2022, *10*, 36429– 36463.
2. Su, Y. Research on Website Phishing Detection Based on LSTM RNN. In Proceedings of the 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chongqing, China, 12–14 June 2020; pp. 284–288.
3. What Is Deep Learning and How Does It Work? Search Enterprise AI. 2022. Available online:

https:/[/www.techtarget.com/se](http://www.techtarget.com/searchenterpri)a[rchenterpri](http://www.techtarget.com/searchenterpri)

seai/definition/deep-learning-deep-neural- network.

1. Saurabh Saoji. Phishing detection system using visual cryptography, 03 2015.
2. S. Patil and S. Dhage. A methodical overview on phishing detection along with an organized way to construct an anti- phishing framework. In 2019 5th International Conference on Advanced Computing Communication Systems (ICACCS), pages 588– 593, 2019.
3. Phishing| Phishing Techniques. Phishing.org. 2022. Available online: https:/[/www.phishing.org/phishing](http://www.phishing.org/phishing-)- techniques
4. https:/[/www.scienc](http://www.sciencedirect.com/scien)e[direct.com/scien](http://www.sciencedirect.com/scien) ce/article/pii/S2665917422001106
5. Grover, R. Deep Learning- Overview, Practical Examples, Popular Algorithms|Analytics Steps. Analyticssteps.com. 2022. Available online: https:/[/www.analyticssteps.com/blogs/dee](http://www.analyticssteps.com/blogs/dee) p-learning-overview-practical-examples- popular-algorithms.
6. Malicious Url Recognition and Detection Using Attention-Based CNN- LSTM-KSII Transactions on Internet and Information Systems (TIIS)|Korea Science. Available online: