# **INTRODUCTION**

## Information security is a significant aspect of communication, which involves safeguarding data and information from unauthorized access and tampering. The protection of sensitive information ensures that it remains confidential and that it cannot be altered by unauthorized users. The secure transmission of data from one entity to another is crucial, and the rise in the number of information attacks over the last few decades highlights the importance of implementing effective information security measures [1]. Cybercriminals attempt to capture valuable information for their own purposes. The security of an organization's information depends on various types of data that the organization possesses. As the world increasingly becomes connected through the internet, it has become necessary for businesses to migrate onto the internet. This has increase the number of buying and selling on the internet. The introduction of covid-19 led to an increase in online transactions due to the fact that countries all over the world went into quarantine. This made businesses and jobs operate remotely through the internet, therefore, there was a drastic increase in the number of phishing attacks on organizations and individuals [2].

## Phishing attacks have become a major threat to online security, causing significant financial loss and reputational damage. These attacks typically involve the use of fraudulent emails, websites, or other forms of electronic communication to trick users into revealing sensitive information, such as passwords, credit card numbers, or social security numbers. The consequences of phishing attacks can be severe, ranging from identity theft to financial loss and reputational damage. To address the growing problem of phishing attacks, researchers and security professionals have developed a range of methods for detecting and preventing these attacks. Machine learning algorithms have emerged as a promising approach to phishing detection, as they can analyze large volumes of data to identify patterns and anomalies that may be indicative of phishing activity (Alshehri et al., 2020; Khan et al., 2021).

## In this thesis, we propose an enhanced random forest model for detecting phishing attacks. Random forest is a popular machine learning algorithm that is well-suited for classification tasks (Breiman, 2001). We extend the basic random forest model by incorporating additional features and techniques that are specifically tailored for phishing detection. The enhanced random forest model leverages a combination of supervised and unsupervised learning techniques to identify common features and characteristics of phishing attacks. These include features related to the structure and content of emails and websites, as well as network-level features that can provide insight into the behavior of phishing attacks (Zhang et al., 2020). We also incorporate several techniques for feature selection and hyperparameter tuning to optimize the performance of our model. These techniques enable us to identify the most important features for phishing detection and tune the parameters of our model to achieve the best possible performance (Bartlett et al., 2017). To evaluate the effectiveness of our enhanced random forest model, we train and test our model using several datasets of both legitimate and phishing electronic communications. We use a range of performance metrics to assess the accuracy, precision, and recall of our model, and compare its performance to other state-of-the-art phishing detection methods (Alzahrani et al., 2019; Singh et al., 2021).

## Our research aims to contribute to the ongoing efforts to improve online security by proposing an enhanced random forest model for phishing attack detection. By leveraging machine learning techniques and incorporating additional features and techniques, we hope to improve the accuracy and effectiveness of phishing detection and help protect users from the harmful effects of phishing attacks. In summary, this thesis proposes an enhanced random forest model for phishing attack detection. By extending the basic random forest model and incorporating additional features and techniques, we aim to improve the accuracy and effectiveness of phishing detection and contribute to the ongoing efforts to improve online security.

## 

PROBLEM STATEMENT

Phishing attacks are a serious threat to online security and traditional approaches to detection may not be effective against new and evolving techniques. Machine learning algorithms, such as random forest, have shown promise in detecting phishing attacks. However, the performance of traditional random forest models can be limited by their inability to handle imbalanced data and to capture complex relationships between features. Therefore, this research aims to propose an enhanced random forest model for phishing attack detection that addresses these limitations and improves the accuracy and effectiveness of phishing detection.

## .

# **REFERENCES**

1. Li, M., Nazir, S., Khan, H. U., Shahzad, S., & Amin, R. (2020). Modelling features-based birthmarks for security of end-to-end communication system. *Security and Communication Networks*, *2020*, 1-9.
2. Subasi, A., Molah, E., Almkallawi, F., & Chaudhery, T. J. (2017, November). Intelligent phishing website detection using random forest classifier. In 2017 International conference on electrical and computing technologies and applications (ICECTA) (pp. 1-5). IEEE.
3. Muppavarapu, V., Rajendran, A., & Vasudevan, S. K. (2018). Phishing detection using RDF and random forests. Int. Arab J. Inf. Technol., 15(5), 817-824.
4. Alshehri, A., Traore, I., & Woungang, I. (2020). Machine learning techniques for phishing detection: A comprehensive review. Journal of Network and Computer Applications, 154, 102470.
5. Alzahrani, A. I., AlGhamdi, F. S., Almulhim, M. A., & Zaidan, B. B. (2019). Phishing detection techniques: A review. IEEE Access, 7, 20491-20512.
6. Bartlett, M. S., et al. (2017). Simple, interpretable models for high-accuracy phishing detection. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1857-1866).
7. Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.
8. Khan, N., et al. (2021). A comparative study of machine learning-based phishing detection techniques. Journal