MSc in Cybersecurity: Privacy and Trust

Analyzing the Effectives of Machine Learning in Detecting DNS Tunnelling

Mosope Adefuye
Supervisor: Dr. James Egan

1. Introduction

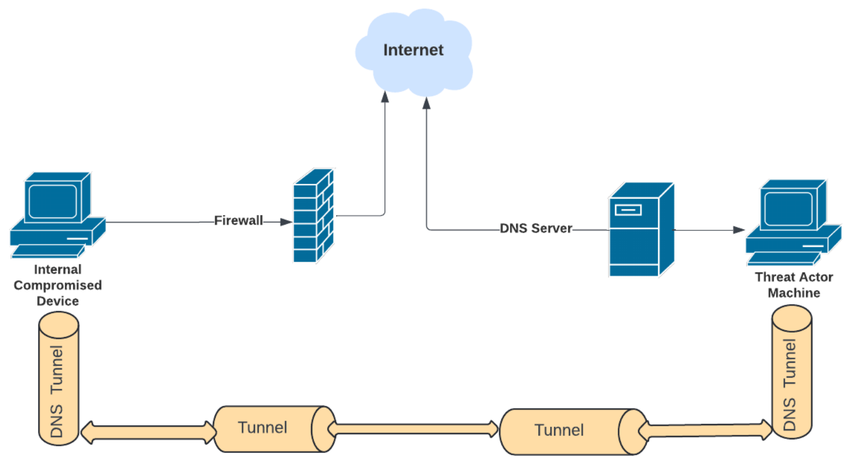
The Domain Name System (DNS), often referred to as the internet's phonebook, facilitates the translation of hostnames into corresponding IP addresses.

Due to the essential role and necessity for organisations to have their domain accessible on the internet, the Domain Name System (DNS), which functions on port 53, is consistently permitted on the firewall, intrusion detection system (IDS), and intrusion prevention system (IPS) by the system/network administrator.

Threat actors often try to exploit the DNS protocol to conceal harmful traffic and avoid being detected by firewalls and other security measures. DNS Tunnel enables assailants to create hidden communication channels and extract confidential information without being noticed.

According to the 2023 International Data Corporation (IDC) threat report, which surveyed 1,000 security experts, the report indicated that 90% of organisations suffer DNS attacks, costing $1.1m each. (Fouchereau, 2023)

This study examines the efficacy of machine learning, namely the K-Nearest Neighbours (KNN) algorithm, in detecting DNS Tunnels. It evaluates the effectiveness of the model in comparison to other models. Furthermore, it assesses if the query size of the DNS request should be considered for the purpose of detecting DNS Tunnelling.



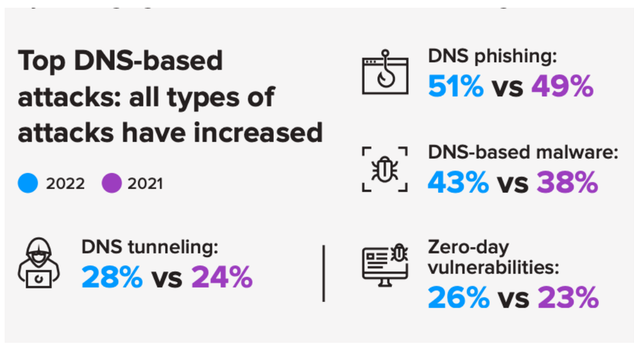
Lab setup

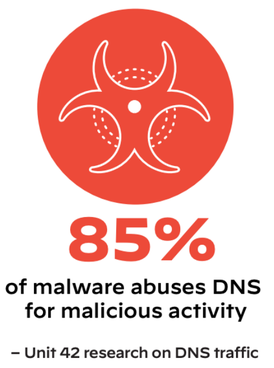
2. Research Questions

- How effective are machine learning algorithms (K-Nearest Neighbours (KNN), Random Forest, XGBoost, Logistic Regression, and Decision Tree) at detecting DNS Tunnels compared to a baseline detection rate?
- How does the selected models perform in terms of accuracy, precision, recall, and F1-score?
- Is the query size of a DNS request a factor in detecting DNS tunnelling?

Hypothesis

- H1: There is a difference between the query size of normal and tunnelled DNS queries



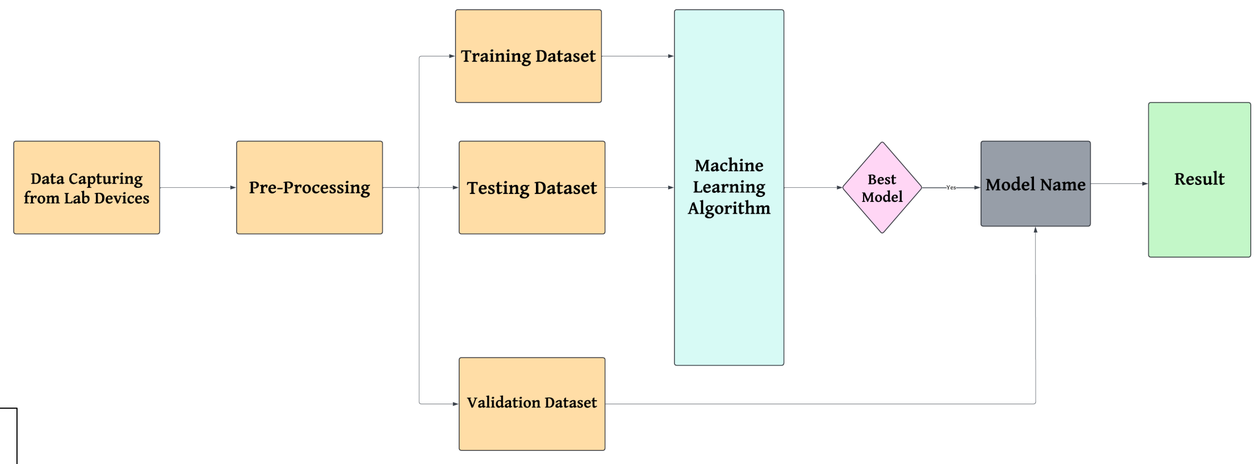


(Fouchereau, 2023)

3. Literature Review

- Packet inspection can offer dependable intrusion detection, but it may lack scalability for extensive networks. (Aiello, Mongelli and Papaleo, 2015)
- Conventional techniques such as Firewalls, Intrusion Detection Systems (IDS), Traffic Analyzers (TA), and Passive DNS Replication have been utilised to detect DNS tunnelling. (Sammour *et al.*, 2018)
- The need for feature analysis, with a specific emphasis on payload analysis and traffic analysis can help in DNS Tunnelling detection. (Sammour *et al.*, 2018)
- The efficiency of the DPI system is highly dependent on possessing prior knowledge of regular traffic data, where the absence of abnormalities within the training data is of utmost importance. (Song *et al.*, 2020)
- Employing both Convolutional Neural Networks (CNNs) and Random Forests (RFs), it becomes possible to capture multiple facets of DNS traffic patterns, hence enhancing the accuracy of identification. (Lambion *et al.*, 2020)
- In contrast to black-box machine learning methods, the security score approach offers explicit reasons for identified traffic, hence improving transparency for cybersecurity operators. (Deri and Fusco, 2021)

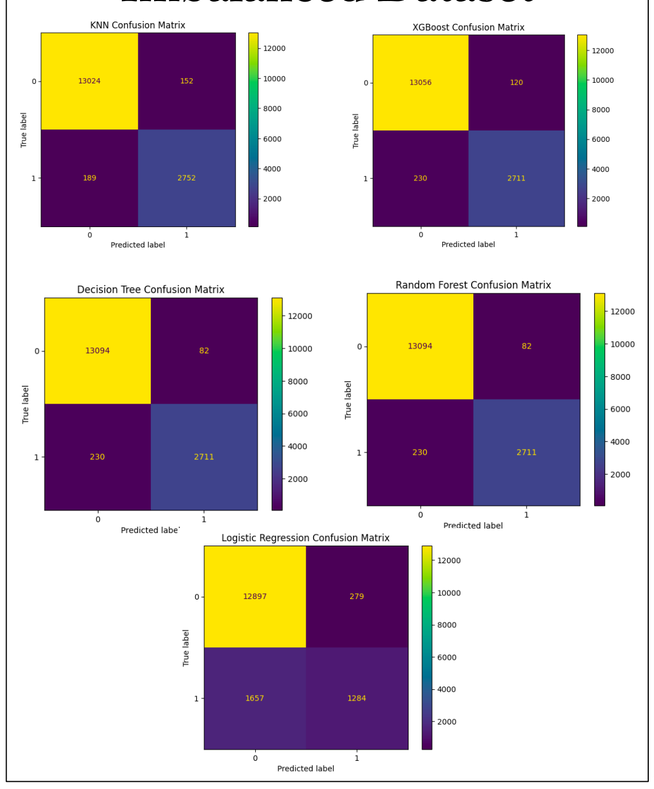
4. Methodology



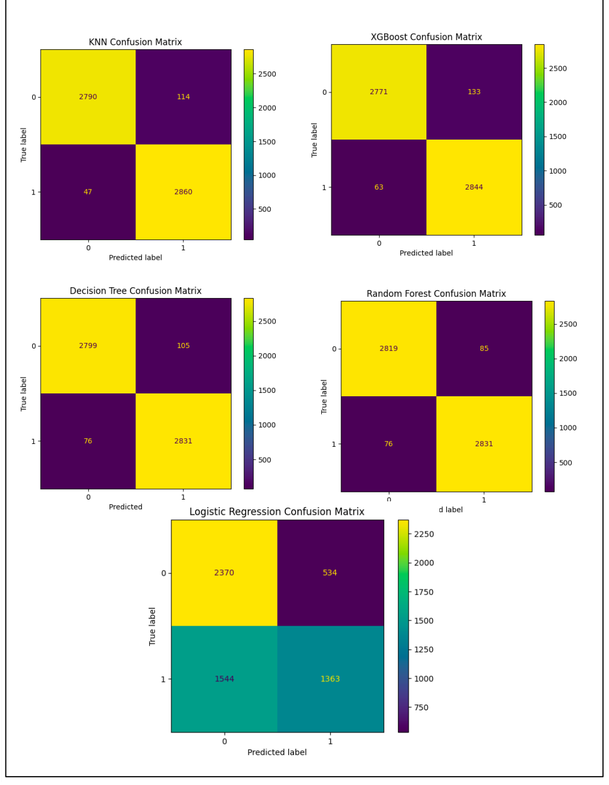
5. Early Indications

- Query size is a factor to consider when detecting a DNS tunneled traffic.
- KNN model tend to have more stability compared with other models as its detection rate remained almost the same in either of the data class balance scenarios

Imbalanced Dataset

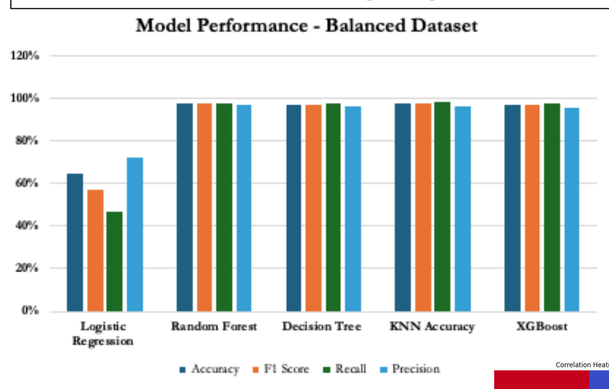


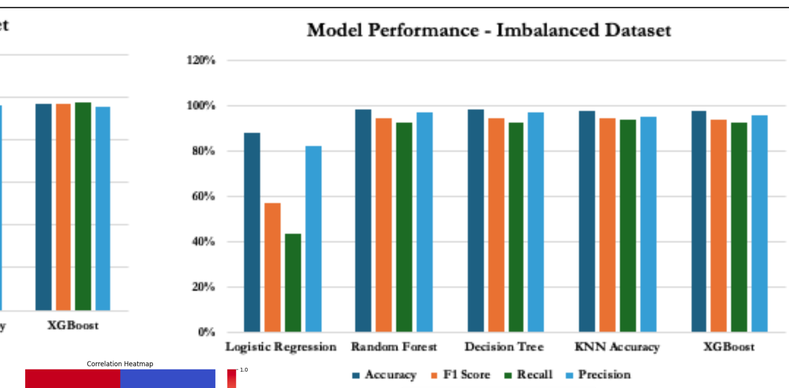
Balanced Dataset

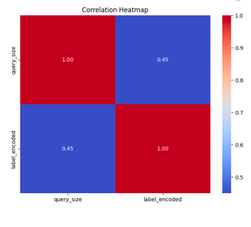


6. NEXT STEPS

- Research on any other variables in the DNS packet that can indicate the presence of a DNS tunnel ?
- Should the roundtrip time of a packet be considered while detecting a DNS tunnel?
- Research if an exponential increase in packet volume within a time frame an indication of an ongoing DNS Tunnel







References

Technology

WIRESHARK · ubuntu · kaggle · scikit learn · python · pandas