|  |
| --- |
| text |
| **DNS Exfiltration attacks with Stateful traffic analysis**  Written by Nasrullah Rami Hamad Under the assistance of Dr. Fawaz Khasawneh |
| |  |  |  | | --- | --- | --- | | Al Hussein Technical University, Amman ,Jordan  21110510@htu.edu.jo, fawaz.khasawneh@htu.edu.jo | 10/8/24 | Computing research project | |

Contents

[1. Introduction 2](#_Toc174219079)

[2. Related Work 4](#_Toc174219080)

[3. Data collection and Description 6](#_Toc174219081)

[3.1 Primary Data 6](#_Toc174219082)

[3.2 Secondary Data 7](#_Toc174219083)

[4. Research Approach and Methodologies 11](#_Toc174219084)

[4.1 Onion Model 11](#_Toc174219085)

[4.2 Research Methodology 17](#_Toc174219086)

[5 Results and Discussion 21](#_Toc174219087)

[6 Conclusion and Recommendations 27](#_Toc174219088)

[7 Reflections 28](#_Toc174219089)

[7.1 Selected Research Methodology 28](#_Toc174219090)

[7.2 Alternative Research Methodologies 29](#_Toc174219091)

[7.3 Recommended Actions and Future Considerations 31](#_Toc174219092)

[7.4 Recommended Methodology 32](#_Toc174219093)

[8 References: 34](#_Toc174219094)

**Abstract**

**This document will provide a throughout research about the DNS exfiltration attack and its analysis on the Stateful DNS traffic. It will go through all the steps required to find a solution for this attack and to collect information on it. This research will help reach a conclusion for the problem via conducting a survey and performing machine learning models on historical data collected from different institutions. Resulting in recommendations for organizations and future works for such research topic.**

**Keywords:** DNS-Exfiltration attack, Big Data, Research

# Introduction

The term "Big Data" has become popular in our evolving age, it changed how data and information is collected from enterprises, stored and analyzed. Moreover, big data means that information has become larger and more continuous and diverse in terms of type. So that the 3Vs were created, volume, variety, and velocity. These three Vs are the core of big data. Evolved technologies and tools are needed to analyze such data and store it effectively.

After conducting experiments on big data, it was seen that the commonly used traditional tools that analyze the normal data isn’t being effective on big data which is why the need for more enhanced tools and technologies are needed to analyze big data. Organizations use many advanced analytical tools to analyze the big data efficiently, and with low effort and time.

Organizations employ various advanced techniques and technologies. Apache Hadoop, a robust open-source framework, includes processing and storage of massive datasets. Apache Spark, another open-source framework, enhances processing speed and efficiency through in-memory computing capabilities, which Hadoop lacks. NoSQL databases such as Cassandra and MongoDB offer the flexibility needed to store unstructured data.

Advanced analytics tools like Tableau and Power BI enable businesses to effectively analyze and comprehend big data. Artificial intelligence and machine learning algorithms are necessary for predictive analytics, uncovering patterns and trends that traditional methods might overlook or ignore.

The implementation of big data technologies significantly enhances operational efficiency across multiple sectors. In healthcare, predictive analytics applied to patient data can optimize treatment plans and resource allocation. In the financial sector, fraud detection algorithms instantly analyze vast transaction datasets to mitigate risks. Manufacturing companies use insights to optimize production processes, reduce downtime and increase the overall efficiency.

**The research problem:**

Mitigating DNS-exfiltration attacks for organizations that deploy Stateful Domain Name Services (DNS) for its devices. DNS-exfiltration attacks take advantage of the DNS protocol inside devices to steal and transfer sensitive data from a protected network to an external server. By encoding the data into DNS queries which are frequently overlooked by the traditional security measures that organizations deploy. Detecting these attacks and preventing them from stealing sensitive data can be quite challenging because DNS is a necessary service in organizations. So, organizations can’t deal with tracking the malicious or benign traffic without disrupting the day-to-day operations inside the organization.

**Research question:**

The question that the DNS-Exfiltration for Stateful DNS traffic Research will cover, starting with working on the dataset, to understanding the problem and the research carefully is:

***Based on the preprocessing and analysis of the features of the DNS-Exfiltration dataset. Can the deployed ML models predict the malicious attacks out of the benign traffic? While deciding whether the attacks can happen or not?***

**Research Objectives**:

these will be the objectives that our research will come across until the end of it. It will include the steps of how our research has developed through its publishment.

1. **Finding a data set that studies DNS Exfiltration attacks.**
2. **Performing preprocessing steps and feature engineering methods for the data set, to involve only the most important features into the modeling phase.**
3. **Starting the modeling phase and studying the performance of the models used and comparing them to find the best model to use for future uses.**
4. **Conducting a survey (Primary data) for the research problem, to get feedback and notices on how it is explained and understood from the public, and its danger on the organizations, and plotting the survey results.**

**What the document provides:**

This document will include the many steps of the research regarding the DNS-Exfiltration attacks in the Stateful DNS traffic. It will include several levels that will generalize and analyze the research from its roots of the related work such as the Literature Reviews that I read, to the conclusion of Results and Reflections on the research topic as a whole. Starting from Section 1 which included the introduction to the whole research. It gave the overview of the problem and its objectives. Section 2 includes any of the works related to my research topic. Which combines the Literature Reviews that were read and analyzed in order to gain a huge in depth comprehension of the topic that I am researching. Section 3 however, includes the collection of the data, with its primary and secondary functions. Which includes the survey that will be inducted as a primary source of data. For secondary, it will include a python code with its models that analyzes and predicts whether the attack will be benign or malicious, and if the attack was malicious, it will be separated between light attacks and heavy attacks. For section 4, numerous research methodologies will be studied, and the Onion Model will be analyzed, to find the research methodology road that the research went through to finally conclude in a fruitful result. Section 5 will include the results of each data collection method that was introduced in section 3. This section will include plots and interpretations of the modeling results, and the survey answers. Section 6 will be the conclusion for the document, and recommendations regarding the research topic will be included. Section 7 will be the reflection part of the document; it will serve as the end of the document and will include my reflection to the entire ladder of steps for the research conducted.

# Related Work

**Archival research is one of the strategies that was used to collect and gather data to further understand the DNS-exfiltration attacks and help mitigate their risks. Here are some of the research papers (Conference & Journal) that were used to fully understand this problem further.**

DNS-Based Data Ex-filtration is one of the threats that target the ever-evolving landscape of Cybersecurity and data. It is a method used by threat actors to steal and transfer data and information from networks to the attacker’s hands. It uses the DNS, which is involved in translating IP addresses into domain names and vice versa, to reach the data inside IP addresses and accounts of victims. The attackers cloak themselves as a normal query, and search for other queries with passwords and usernames, until they reach the mother server, which will lead them to the user’s credentials. This attack poses a significant threat for the users and the companies in general, because most organizations rely on cloud storage and digital infrastructure to store their databases. Which can be targeted by this cyber-attack. Many organizations have integrated AI and other technologies in order to battle this security threat. By training models on historical data based on attacks and enhancing them to catch such behaviors.

**In J. Diao, B et al., (2022).** This paper studied the many ways of integrating AI into defending against such models, they noted that the AI models weren’t as effective because they couldn’t get a hold of many data regarding that attack. So, they implemented their model which was called TTP, they also introduced a new decision-making technique that converts human-made decision making into a dynamic machine decision making system. It is used against the DNS-based attacks because they can identify the attack easily and accurately. Which can help use AI-powered methods.

**In Mahdavifar et al., (2019).** The paper studied DNS and referenced how it’s the most common way to steal information and data from organizations and users, and to counter such attacks they use firewalls to protect their information. But they were also infiltrated by using encoded data to a compromised server controlled by the attackers. So, they developed a hybrid approach that uses features, that uses stateless and stateful features. This approach was then integrated into an AI model that used a massive data-set generated by exfiltrating various file types and sizes. Then by using preprocessing techniques, the data-set ended up with nearly a million instances. By using ML algorithms, a detection system against DNS-based exfiltration attacks was created.

**In Ahmed, J et al., (2020)**. This paper studied the way the DNS service being the most attacked part of a network, because it can easily bypass the security of firewalls without any deep inspection for the networks and IP addresses. So, to counter that, the paper provided a way to detect these exfiltration attacks over the DNS, what makes this solution unique is that it works on real-time analysis of real DNS traffic. It was used by two organizations, to extract a portion of their data on a 3-day basis. They used a stateless system to counter the DNS messages that included malicious intent. They also contributed in developing a ML algorithm to detect the anomalies in DNS queries using a dataset of domains. By selecting a 14-day data full of DNS traffic. They used the ML model on the two organizations and injected them with three million malicious queries that were sent via the DNS server. And they identified them with a good accuracy. Then, they compared the results of the two methods. They informed us that the tools they used are available for the public and further research.

**In Bubnov, Y.et al., (2018).** The paper describes the DNS tunneling attacks and that attack organizations and users and steals their information and credentials. It notes the methods organizations use to prevent such attacks as inefficient and useless. But the paper has notified the readers that they developed a multi-labeled model that is trained on previous DNS traffic between users. It works on classifying the tunneling strategies that are popularly used on a wide track of attacks. Which can identify such attacks between users. The paper studied instances of usage of the method. And included analysis of accuracy for the model.

**In Shashank Biradar et al., (2019).** This paper studied the DNS cache poisoning attacks that target users´ information. And established a machine learning model to counter it. The features they selected were essential DNS packet components. Such as the trigger of the DNS protocol, time-related features, location-related features, and more. The paper noted that DNS works by mapping an IP with a domain name, which leads users to the right websites and data. Then, the paper suggested that its model works on comprising training with DNS traffic data on some DNS security weaknesses, such as DNS-based malware, DNS-amplification, DNS tunneling, and more. This model works on evaluating the small-scale networks (experimental environments) and the large-scale networks (real environments).

**In Izabela Savić, et al (2024).** This paper started by showing how DNS is the key to accessing all websites by users rather than remembering the IP address of a website. Then, it showed that DNS exfiltration methods are appearing more frequently and more dangerously than ever, this goes to the lack of detection and attention to such attacks. These attacks could lead to organizations and user data stolen from them through the exfiltration. Then the paper mentioned that Machine learning is one of the counters for such attacks, since it works on detecting the attacks and prevent them. But it lacks good accuracy since many real DNS traffic packets are also trained within the machine learning models, which can lower the accuracy and reliability of such models. But the paper then explained that some machine learning algorithms that were based on voting ensemble methods were successful in attack detection. This method increases the accuracy and provides levels of prevention and protection against such attacks.

**In Saha et al., (2023).** This paper mentioned Cyberattacks that are emerging in fast speeds, and the way they are performed. With the numerous tools and ways that attackers use, from phishing, malware, ransomware, to DNS exfiltration. The paper noted that DNS is commonly attacked due to its vulnerability and because its attacks go unnoticed. The paper went on to talk about a solution to such attacks, which is using HTTPS websites. Because the DNS would then travel across secure traffic, this solution is known by DNS over HTTPS (DoH). But attackers have started using encrypted IP addresses to hack and steal the information from the users and the organizations. So, the Canadian institute of cybersecurity started working on a solution for this, and created a machine learning method based on ensemble feature selection techniques to classify the traffic of the DoH. The paper then studied several machine learning methodologies and fed and trained them with datasets extracted from the traffic. And according to their analysis, the ensemble feature selection model had the best performance of the other models used.

Many of these Literature reviews didn’t include a massive dataset to analyze them with ML models. This is the gap that my research will cover while implementing feature selections and oversampling methods. This will give a whole new experiment and a result to this topic. Which will evolve this study for further inquiries.

# Data collection and Description

The main difference between primary data and secondary data is the primary data was collected firsthand by the researcher (me), and it consisted of information that I needed to enhance the research process, I used surveys for the primary data collection. Secondary data refers to the data that was collected by the researcher (me), but wasn’t collected firsthand by him, such as academic journals, archival researches, datasets used for experiments. For my secondary data, I used a dataset from sensors that detected stateful traffic inside a cyber security organization. To help find ways to mitigate the DNS-exfiltration attacks for all organizations.

3.1 Primary Data

Describe your primary data (it could include figures and tables).

My primary data collection method, will be a survey conducted for 50 people that are experts in the Cyber Security field. Every question inside the survey will be a quantitative (Close-ended) question. It will be very easy and simple to understand the questions which will be clear for the individuals. The results will be numerical and will be interpreted in form of plots and graphs that will be comprehensible for the readers and researchers to understand.

Describe your experience in collecting your primary data. Include any issues you faced in this process.

Finding the right individuals to submit their answers to the survey was difficult, because not many people were familiar with the DNS-Exfiltration topic. Since it is a new attacking format used to steal sensitive data from individuals and organizations. But I was able to find the right number of answers for the survey in the end. The selection of questions was also difficult, since I needed to find the right words to make the questions clear and understandable for the individuals so they submit accurate answers regarding the topic.

Justify your choice.

I chose a survey as a primary method of collecting data, because I have used it before in other researching topics. And it was effective in gaining the right number of results and graphs. Which happened to be the result for this type of research I’m working with this time. It was significantly effective in this research. Since the results and plots were understandable and convincing for the majority of people.

Describe the merits and limits of primary data.

There has been many limits and merits in using surveys as the primary source of the data for this research. Some of them are:

Merits:

* They’re easy to use and make. Using Google Forms as a platform for individuals to use and write their answers inside.
* It is time-efficient, and can give results in a timely manner.
* It presents plots and graphs for every question as a result. Which can be understood by people with no expertise in the topic.

Limits:

* Such method can produce good results, but it can be time consuming to answer long and a huge number of questions. So, some people might avoid answering the survey.
* Finding the right clear words for questions might need time, because the results need to be accurate, so I need the individuals to understand each question correctly before answering.
* The analysis of plots can be time consuming, because some of the plots may not be in the same shape that was desired from the question. This might need to use other analytical applications to view the plots in a different format.

3.2 Secondary Data

Describe your secondary data (it could include figures and tables).

The secondary research used is a python code that uses a dataset that includes DNS-Exfiltration attacks, with a classification target, (is\_attack) This target holds 2 classes, attack and benign. The dataset studies the stateful behavior of the DNS traffic. It holds 262,105 records and 28 different features, each of them is necessary to study the exfiltration attacks for the DNS. The independent features are the following:

1. **rr\_type:**resource record type. For example, A, AAAA, NS, MX, TXT, etc. Note: using SRV, TXT, or NULL records in DNS queries can be considered suspicious traffic because legitimate clients don't typically resolve domain names to non-address resource records.
2. **rr\_count:**the number of DNS resource records in each section of the DNS message. The sections of the DNS message are 1) question, 2) answer, 3) authority, and 4) additional.
3. **rr\_name\_length:**the length of the resource record name.
4. **rr\_name\_entropy:**referstohow much the resource record name is readable.
5. **rr\_type\_frequency:**for a certain domain, rr\_type\_frequency equals the number of packets of a specific resource record type over the total number of DNS packets for that domain.
6. **rr:**the rate of A and AAAA records for a domain during a specific period.
7. **distinct\_ns:**the total number of name servers (NS) that were resolved in the DNS database. i.e., total count of distinct NS records.
8. **a\_records:**total count of IP addresses resolved in the DNS database.
9. **distinct\_country:**“Distinct country names for a given domain in window τ” [2].
10. **unique\_ASN:**Unique autonomous system number (ASN) in window τ.
11. **unique\_ttl:**“Distinct TTL values in window τ” [2].
12. **distinct\_ip:** “Distinct IP values for a given domain in window τ” [2].
13. **distinct\_domains:** “Distinct domains that share the same IP address that resolved to a given domain in window τ” [2].
14. **reverse\_DNS:**“Reverse DNS query results for a given domain in window τ” [2]
15. **ttl\_mean:**the mean of TTLs (time to live) in window τ.
16. **ttl\_variance:** the variance of TTLs in window τ.
17. **A\_frequency:** indicates how often A\_records are encountered in the DNS traffic
18. **NS\_frequency:** indicates how often NS records are encountered in the DNS traffic
19. **CNAME\_frequency:** indicates how often CNAME records appear in the DNS queries
20. **SOA\_frequency:** indicates how often SOA records are encountered in the DNS traffic
21. **NULL\_frequency:** indicates how often NULL records are in the DNS traffic
22. **PTR\_frequency:** indicates how often reverse lookups are performed
23. **HINFO\_frequency:** indicates how often the HINFO records appear in the DNS traffic
24. **MX\_frequency:** indicates how often mail server information appears in the DNS traffic
25. **TXT\_frequency:** indicates how often text information is included in the DNS traffic
26. **AAAA\_frequency:** indicates how often IPv6 addresses appear in the DNS traffic
27. **SRV\_frequency:** indicates how often server located information is queried in the DNS traffic
28. **OPT\_frequency:** The frequency of OPT records inside the DNS traffic

And the target feature **is\_attack:** which determines whether the traffic could have an attack or not.

These features will be used to detect the DNS exfiltration attacks by training machine learning models on them. So that we can understand and prevent such attacks from occurring in organizations and other institutions.

By using the command DF.info(), I showed all the datatypes of the features and their records.

It is shown that the dataset holds 262,105 records for all 28 features. Which indicates that there weren’t any NULL or missing values.

The datatypes of the features as shown are Integers, Floats, and Objects. To turn the objects into numerical values to train the models, we will use encoder techniques such as the Label Encoder.

Some of these features may not be scaled enough to provide a good model prediction. So, we may need to use a normalization technique such as the standard scaler.

The target variable (Is attack) must be balanced (all classes are in the same range) in order to prevent biased insights. We can use SMOTE as a balancing technique for the target so that it can provide a good read.

The feature amounts is high, and it may not provide significant insights, so we might use feature selection methods to eliminate useless and unnecessary features from the dataset.

Justify your choice.

The purpose of machine learning models to analyze and prevent this type of attack can vary for many reasons. Machine learning models can deal and handle complex patterns, and this type of attack may be difficult to detect using traditional rule-based systems. ML models can also analyze the behavior of such patterns in real-time to identify any anomalies inside the traffic. What makes Machine learning models a good pick, is their handling of high data volumes, as mentioned before, the dataset holds 262105 records which can take forever to deal with using rule-based systems.

Machine learning models are open to generalization of new attack patterns and previously unseen attacks. Which can help with detecting the rare DNS exfiltration attacks. While also providing insights to the organization and the users of such models by giving metrics and prediction accuracy for the attack pattern.

Such models are efficient for real-time analysis and are resourceful when it comes to monitoring the data, which can allow security teams to focus on other tasks. These models have the ability to engineer the features and change them entirely, this might be useful if the need for feature selection is necessary.

In conclusion, machine learning models can study, detect, and interpret the results of the DNS exfiltration attacks on the stateful traffic, while analyzing them in real-time. So that organizations can be guaranteed protection and better prevention for such attacks.

Describe the merits and limits of secondary data.

Working with machine learning models and using them as a secondary source of research is a strong valid option, since it can be a good source of non-human interference. A good model can be recommended to be used for some organizations that have poor security. There have been some merits and limits when it comes to the ML models. Such as:

Merits:

1. Machine learning models are able to understand complex patterns inside the datasets. Which makes it the best at handling and finding correlations between the dataset features that could be missed by a traditional method.
2. It has the ability to handle huge amounts of data that come with high volume (Big data). Which can help by analyzing them efficiently, and giving a real-time analysis for them.
3. It enables organizations of predictive abilities, by providing insights and predictions based on future events by training the models on historical data collected by the organization or anywhere else.

Limits:

1. The data collected by organizations (Raw data) must be preprocessed and handled effectively in order to be able to train it on machine learning models, this can be time consuming which varies on the task and the data volume.
2. Machine models trained on high volume data can be resourceful, and will require high computational devices to work effectively. Making it computationally costing. It may also need an infrastructure to support the demands of such machine learning models.
3. Data privacy and continuous maintenance for the models and data are two big problems that are encountered when dealing with such models. To remain accurate with good performances, ML models require continuous maintenance to provide insights that counter the data evolution and changing. The data collected can have biases or privacy concerns that may deal some problems inside the organizations before working with them.

# Research Approach and Methodologies

This section includes the Onion model which is a layered concept with each layer holding several methodologies a research project can go through. This section will hold every layer, with every methodology discussed, While getting in depth with the methodologies I chose for the researching project for the DNS-Exfiltration attacks with the Stateful analysis. In the end, a research methodology graph will also be discussed and will be critically evaluated on how my project was made while evaluating the setup, the conduction, and the end of the project.

### Onion Model

The onion model is a framework designed by Saunders, Lewi, and Thomhill. For the purpose of support for researchers to approach their studies in a systematic way. It presents a structure that guides them through stages, each stage of them holds multiple methodologies that researchers can choose from to further conduct their research into a successful one.

It is a layered concept that enables six different layers to choose from, each layer having several different methodologies that each research would have to go through, this structure provides a sequential nature for the research design, and encourages the researchers to choose multiple methodologies to further enhance their research before progressing to the next layer.

This model is known for its flexibility and clarity, since it provides a wide selection of clear and well-structured methodologies and designs to choose from, to ensure that researchers find the best fit for their needs and tasks. It also holds a mean of integration between layers, so if a researcher chooses a methodology from a specific layer, it will affect the choice in the next layers. This helps with the consistency of the research design for the researchers.

This model is used for planning projects, since researchers can use it to outline their designs and align their necessary researching elements, and it allows researchers to make efficient and effective decisions when it comes to planning and identifying the next step in their research. It can also be used as an educational tool to help students grasp and get hold of the complex steps in a researcher’s job while understanding the relationship between its different stages and methodologies.

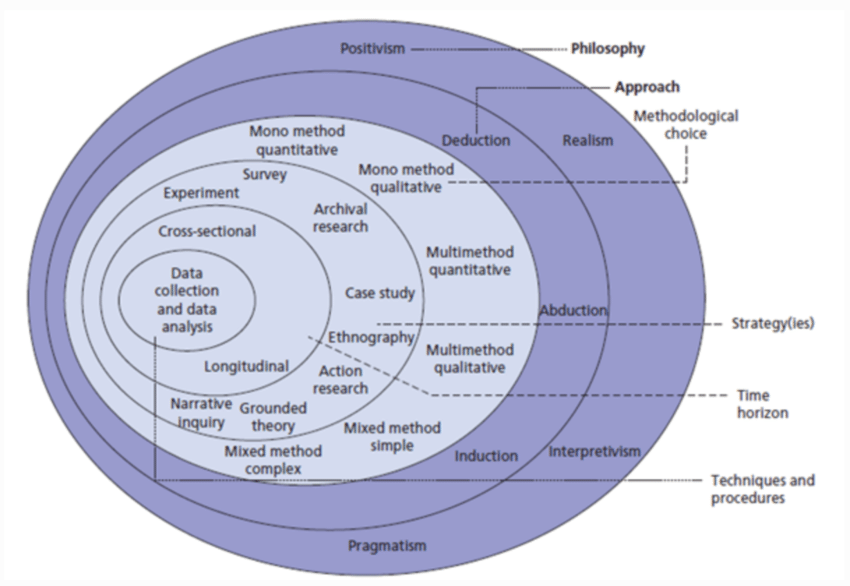


Figure 1: The structure of the Onion Research Model [1].

* + 1. **Philosophy**

This layer studies the reality of the assumption that will be researched in later stages. This layer will be the beginning of the influence for the researching approach throughout the entire model. It contains methodologies that study the nature of the assumption whether it is seen as a social experiment or from a certain viewpoint.

1. **Positivism**, which focuses on how the objective seeks to identify the general laws through testing and analysis, focusing on a specific viewpoint.
2. **Interpretivism**, which suggests that the research can be subjective, and will be understood through human and social experiments and interactions.
3. **Realism**, mentions that the focus of the research is a reality whether people agree upon it or not. It can be understood through human understanding.
4. **Pragmatism**, can be a mixture of both positivism and interpretivism, since it can be seen from a certain viewpoint while conducting practical experiments and solutions.

I used **Positivism** in my research about DNS-Exfiltration attacks, because through various steps and levels of testing and analysis, the topic of the research resulted in a single viewpoint by predicting whether the attack happened or no. **Realism** was also a philosophical step that was used during the research, since whether people agree or not, the prediction for the attack was set beforehand. Interpretivism and pragmatism weren’t used because the research was mainly based on the hypothesis during the coding part, and not through the human experiments and interactions.

* + 1. **Theory Development Approach**

This layer contains the strategy that the research will go through, starting from data collection, or observations, or a theory, and ending with either of them.

1. **Deduction** starts with a theory or a hypothesis to end up being tested on observations, data collection methods, and strategies. To provide fixed and guaranteed results on the selected topic.
2. **Induction** starts with an observation based on some patterns that ends up generating theories and assumptions on the selected problem.
3. **Abduction** starts by developing theories and efficient hypotheses on a specific observation, and ends up with a prediction, this can result in inference about a specific problem.

I used the **abduction** theory development approach, since I used machine learning techniques and algorithms to generate predictions on whether the attack happened or no based on massive amounts of research and hypothesis testing, to finally end up with a good inference for the attack. Since I happened to reach predictions and insights at the end of the research, deduction and induction cannot be used in the research, because deduction focuses on guaranteed results, not predictions, and Induction focuses on general results on the selected topic.

* + 1. **Methodological Choice**

This layer involves how data will be collected and in what form, to be later analyzed.

1. **Mono-method quantitative**: This methodology implies that a single data collection technique will be used to collect quantitative results by a certain procedure.
2. **Mono-method qualitative**: This methodology also implies using a single data collection technique to collect qualitative data and results with a specific way.
3. **Multi-method quantitative**: This methodology states that multiple data collection techniques will be used to capture quantitative data based on a system.
4. **Multi-method qualitative**: This methodology presents multiple data collection techniques used to gather qualitative data by a specific procedure.
5. **Mixed-method simple**: This methodology uses both qualitative and quantitative techniques in order to understand the researching problem better, the simple method uses simple amounts of techniques to understand the problem.
6. **Mixed-method complex**: This methodology also uses both qualitative and quantitative techniques to understand and find the best analysis for the researching problem. But it uses more complex techniques to collect data and analyze the research more efficiently.

I used the **Mixed-method simple** for my research problem, I conducted a survey that generated quantitative results because the questions were close-ended throughout the survey. I also did a comprehensive python code that generated qualitative results on whether the attack happened or not using a classification dataset problem. The result was a mixture in both qualitative and quantitative methods, and the methods selected were amounted to two only, so that means it was a simple mixed-method.

* + 1. **Research Strategy**

This layer represents the plan that is going to be taken to gather and analyze data.

1. **Experiment:** These are conducted by applying variable manipulation and relationship between features and other data in a single controlled setting. Field and laboratory experiments are examples on this strategy.
2. **Survey:** By gathering data from individuals for a specific problem, in a form of questionnaires. The questions can be close or open ended. It can be conducted by surveys done online, or by interviews with the individuals themselves.
3. **Archival Research:** Investigating a specific problem by usage of previous historical documents and papers.
4. **Case Study:** Doing a study on one of a number of problems within a real-life aspect. It is often done in documentaries.
5. **Ethnography:** Gathering data from constant observation on an individual’s life, the research is done with agreement from both sides, and it focuses on a specific time-scale.
6. **Action Research:** Involves getting in-depth with a problem through constant planning and reflection. This form of strategy can be used to solve practical problems.
7. **Grounded theory:** This strategy involves creating new theories based on historical data and continual analysis.
8. **Narrative inquiry:** This study uses narrations from individuals about specific problems that are the subject of the research conducted. It can vary from an individual’s point of view, but it involves detailed explanation from the story.

I used three strategies here to collect data, each strategy was used separately, on the same topic, each strategy had a different way of implementation. But the combination of the three strategies allowed me to further strengthen the research efficiency and performance and accuracy. Which resulted in better decision-making and final research performance. I used archival research as a first step in gathering information about this type of DNS attack and to collect data on how cyber security attacks work. I used surveys because they happen to help gather information from individuals about the specific type of DNS attack, which helped gather intel from professionals from universities and the business environment. I also conducted experiments using a python code and a classification dataset, which resulted in predictions on whether the attack would happen or not based on a number of independent features gathered from the sensors of the DNS traffic. The combination of these three strategies made me possible to produce a fully detailed research topic on the DNS-exfiltration attack effectively.

* + 1. **Time Horizons**

This stage refers to the time period that took to gather the data from the strategies.

1. **Cross-sectional**: The data in this stage is collected in a single period of time, at once. For example, surveys, interviews, and grounded theories.

1. **Longitudinal**: The data in this stage however, is collected in an extended period of time. This is due to the amount of the data or how the strategy involves data collection, such as experiments and ethnography.

I used a cross-sectional period of time, since the predictions from the python code, and the survey conducted can have their data collected in a single period of time, not over a timescale.

* + 1. **Techniques and Procedures**

In this stage, the specific techniques for each data collection method are present.

Since I have used three strategies to gather and collect data, each one of them had a way of implementation and a structure of how data was collected. As the following:

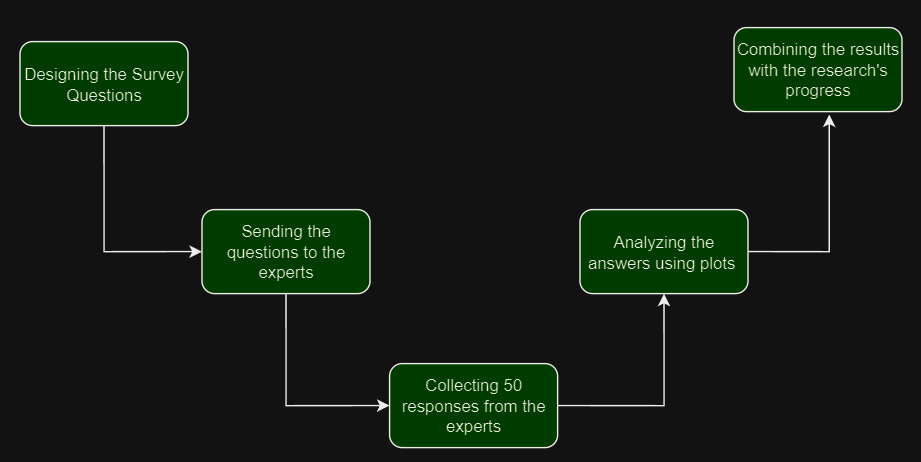
1. **The conducted survey**: The survey conducted consisted of 10 close-ended questions that each had a clear and concise question needed to be answered from professionals to gather the information directly from individuals. Approximately, 50 responses were gathered from cyber security experts using this method, and the analysis was done by Google Docs’ plotting system where each question provided a plot as a result of the question. At the end, all questions were analyzed effectively and the answers were then added to the researching journey.
2. **The Python code**: The code implemented using the Python language had a classification dataset problem that collected data from real sensors inside a cyber security facility on the DNS traffic inside devices. It focused on the Stateful side of the traffic, and the goal was to enhance the quality of the features and the data inside the dataset, to finally provide a prediction on whether the attack would happen or not on the stateful DNS traffic. This strategy was implemented on various ways of preprocessing and analysis of the data:

* **Feature engineering** was used to find ways of either reducing the useless and irrelevant features or data points, or to produce further efficient features that can help in stabilizing the machine learning models and finding good correlations therefore providing with great insights.
* **Encoding** function was used in order to find all the string and object features and transforming them into numerical features so that they can be split and then trained effectively into the machine learning techniques without having to encounter errors while doing the prediction stage.
* A form of **scaling** was used in order to present all the data points more clearly and have them on the same bounds, called Standard Scaler. It was done on all the numerical values after the encoding step happened. This helps with the training phase of the code.
* More **preprocessing** steps such as checking for missing values and handling duplications was made, I checked the data for any outliers inside the features, and made sure to eliminate them completely.
* I used **EDA and plots** for analysis such as Heat maps to have a good look at the correlations between all the independent features and the target variable. So, I can choose a good number of features to start the modeling phase on.
* I used a **balancing** technique (SMOTE) that creates synthetic samples to provide a good balance between both classes of the target (Benign & Attack). It helped with the overall state of the dataset, so that the accuracy measure can be used to provide further insights into the model.
* I conducted a **two-way split** into training and testing data for the dataset, on a ratio of 70:30 percent. So that I can fit the training data into the models and compare them with the testing data. To provide clear insights on whether the attack would happen or not. Based on the dataset.

Then after the preprocessing and analysis steps that were conducted above, I started the feature selection process, by selecting three feature selections each from a different type (Wrapper, Filter, and Embedded) I set the number of features selected to 8. Then collected all the features taken from the methods and input them into the modeling process. Which consisted of three machine learning models (XG Boost, Logistic Regression, Random Forest). And three neural network models (Simple RNN, GRU, LSTM). All models were trained on the training data and then tested with the testing data, and then evaluated using four metrics that are commonly used, Accuracy, Precision, Recall, and F1-score. The models generated a great performance with accurate insights, and the results were then taken to improve the researching validation even further.

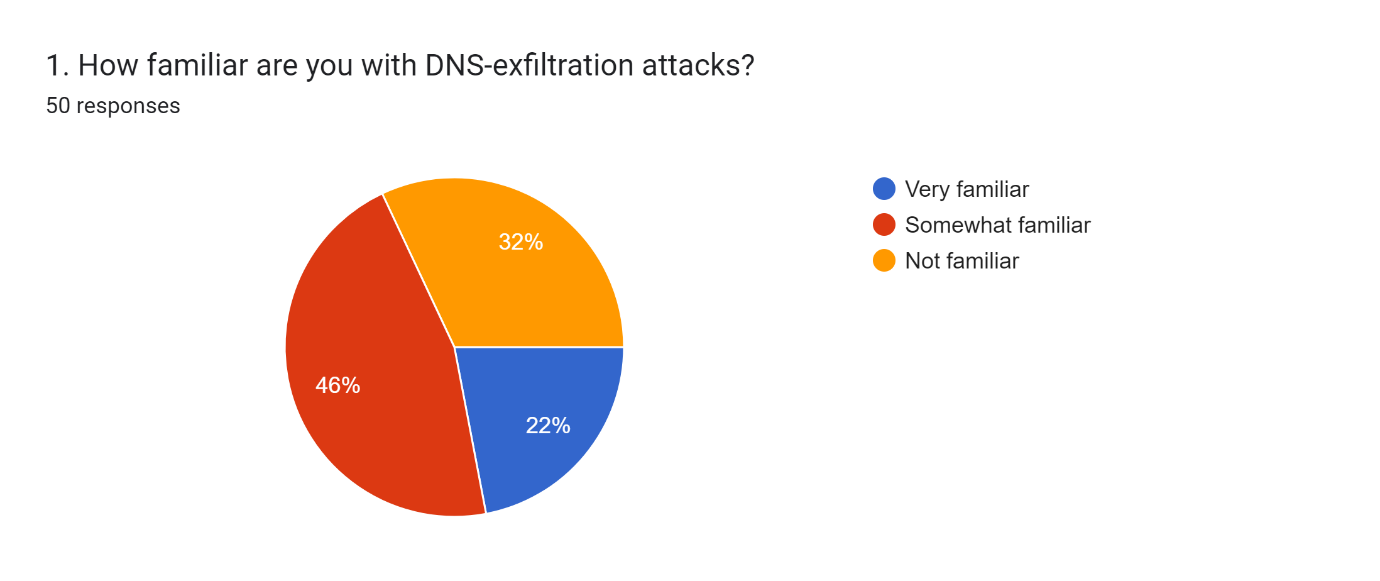
### Research Methodology

I have done two strategies or methods to gather data and integrate it into the research progress to validate the research even further. The survey, and the Python code. The survey conducted had these steps done in its implementation as shown in the Graph:



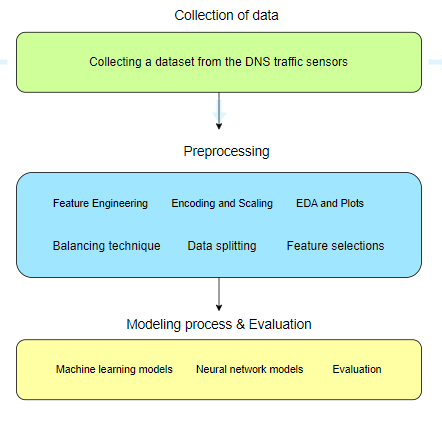
Each of the steps of the survey implementation happened to have a huge influence on the result and research accuracy and validity. The steps had their weaknesses and strengths.

1. **Designing the survey questions**: This step is significantly crucial to the whole survey, because if a question had an unclear language or topic, the answers would be different to what is needed for the research, therefore ending up in total failed research attempt. Clear and concise questions with an understood language that experts can understand and respond to is necessary. The questions were written in a close-ended format, totaling 10 questions throughout the survey. Choosing the right language was challenging at first because they need to be in a professional language that cyber security experts can understand to gather valuable information and results.
2. **Sending the questions to the experts**: This stage had some ups and downs, because finding the right experts was difficult and challenging at the start, but after some consultation from the university. I happened to get in touch with a cyber security institution that helped gather the responses quickly and in an effective manner. The gathering process for the results started then. Choosing the right experts is necessary to gather useful data that can be used to validate the research further.
3. **Collecting 50 responses from the experts**: After finding the right individuals for the survey, it is time to start gathering data and information from them about the DNS-exfiltration attacks. The institution that I was recommended happened to be Cybersecurity related, so the collection of the responses will certainly be detailed and quality assured. The gathering of the responses took some time, but it was worth it because the responses will be efficient with quality and useful information to add to the research progress.
4. **Analyzing the answers using plots**: After the gathering session that took place inside the institute, it was time to analyze the 50 responses, and the best way to do it is by using plots and graphs. Each question had a single plot that viewed the answers from the 50 experts in a clear representation. The plots used were Pie charts because of their flexibility for close-ended question answers, and for its clear structure and easy-to-understand implementation. Example on a plot from the survey:



1. Combining the results with the research’s progress: Finally, after gathering all the results from the experts. They will be combined with other data collection strategy to validate the research’s performance, and to ensure that its robust with detailed evidence. The data collected from this method is going to provide the research with expert opinions on the DNS-Exfiltration attack which will be compared and combined with the results from the other method (Python code).

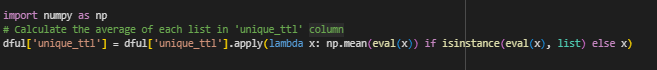
The survey itself has been a great strategy to gather information by, from the help of the cybersecurity institution and the 50 experts in the field. I was able to analyze enough responses needed to have a point of view on how a number of organizations can handle the usage of machine learning in DNS-exfiltration attacks, and if they’re familiar with the attack in the first place. Some organizations would be ready for any attack, and their response time would be short compared to some organizations where the attack would surprise them and potentially damage their data and steal sensitive information from them. While their response time being in days. Some organizations have a well-trained team of experts to monitor traffic and respond effectively to the attacks, while some organizations don’t. It is notable that the integration of machine learning algorithms inside the cybersecurity field was a difficult task, because of its computational power, data volume, quality, and the negative evolution of security attacks.

The second data collection strategy that was used to gather data and further strengthen the progress of the research for the DNS-Exfiltration attacks is a Python code implemented to predict a classification problem of whether an attack was going to occur on the stateful traffic based on a number of independent variables, as shown in the graph:

1. **Collecting a dataset from the DNS traffic sensors**: To start with the second strategy, which is the implementation of the python code, a bunch of traffic-reading sensors were placed over a period of time, to gather data on the stateful DNS traffic, and record them in a structured database (CSV file). The stage is very necessary because the quality of the data and the number of features will be generated from this stage, further steps will update on the dataset collected from this step, and bad collection of data can lead to various errors and misreads in the dataset, which can lead to bad results and therefore a failed researching attempt. The collection of the data was smooth and it was collected in a timely manner, and the data was secure and the ownership was validated.
2. **Feature Engineering**: I implemented some feature engineering steps that benefited the dataset, such as removing some columns that had all their data at value 0.



this helped with keeping the features that can positively benefit and affect the models without interrupting other variables and reducing their correlation with the target. I have changed some of the data inside the “Unique TTL” column:



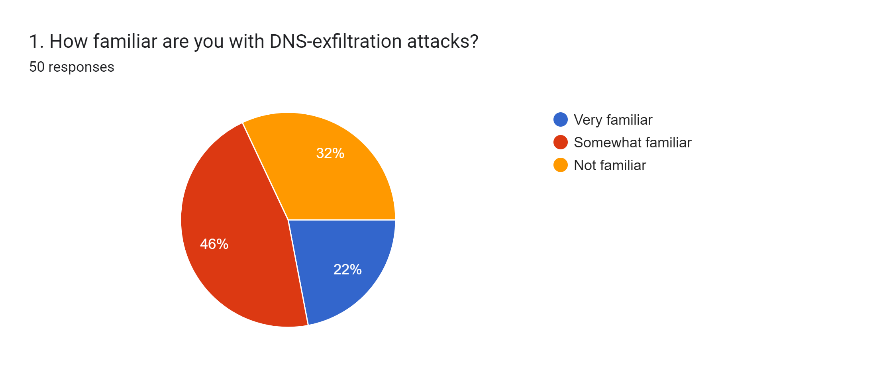
1. **Encoding & Scaling**: These are some of the most essential steps in every code that results in predicting data, because it helps with providing a clean dataset to the models to generate great results and insights. I have used the Label encoder to encode and transform the object variables so they can be numerical, since most machine learning algorithms need the data to be numerical in order to predict the target variable on them. This way, the predictions can be clear and accurate and the machine learning models can be clear of any errors inside the python code. I have used the standard scaler as a form of normalization to set bounds to the data inside the features (ranging between 1 and -1). This step helps with the machine learning model to be trained on data that is on the same range, which helps in providing better results and accurate insights.
2. **EDA and plots**: I have used a heatmap to provide the correlation for each feature with the target variable, this way I can see all the features that are relevant and those that aren’t. And then, we can perform the right technique to get rid of the irrelevant feature, so that the models can predict the data on the features that are relevant therefore providing us with real and significant results. Which I will talk about in later stages.
3. **Balancing technique**: I have used SMOTE as a balancing technique to counter an imbalancing problem that the dataset was suffering from. Imbalancing means that the target variable has a majority class and a minority class, the balancing technique SMOTE works on creating synthetic samples for the minority class based on the other independent variables to reach the same amount as the majority class. This way the data can be equally trained with no bias towards a specific target class, in this case, the majority class. The data will be trained with and without SMOTE.
4. **Data Splitting (Training and testing**): The data is finally preprocessed and ready to be trained, but before that, we will need to split the data into training data which will be fit to the models in order to train them, and testing data, which will be compared with the data that was fit into the model, in order to provide insights and predictions. Normally the X (independent variables) and the Y (Dependent variable) are split into X test, X train, Y test, and Y train. The training data was granted 70% of the data, and the testing data was granted 30%.
5. **Feature selection**: A crucial step to do to further validate the research, is to use feature selection methods, so that some irrelevant features can be eliminated and only the useful relevant features that can affect the prediction of the attack’s occurrence are kept inside the data. I have used three feature selection methods for this problem. Wrapper method: Coefficient correlation, Filter method: Select K best, and Embedded method: Lasso Regression. Each works in a different way to eliminate any irrelevant feature. They were done before and after the balancing to validate whether the features are better with balancing or without. The coefficient correlation FS works on choosing the features with the best correlation to the target. While Select K best FS works on choosing the most suitable number of variables needed for the target. Finally, Lasso regression works on changing the weight of some of the features that fall below a certain threshold to 0, which eliminates the feature completely.
6. **Modeling process**: The modeling process consists of three machine learning models (XG Boost, Logistic regression, and Random Forest), and three neural network models (Simple RNN, LSTM, and GRU). This step is where the prediction stage starts. By applying the training data and fitting them to all the models and comparing them with the testing data, we can finally ensure the predictions done right for all the data. The closer the predicted data is to the actual data, the better the accuracy of the model. The models used are classification models because the problem is a binary classification problem (2 target classes). XG boost works is a boosting algorithm that works on training parallel decision trees to form one conclusion for all the models. Logistic regression works on training the data on the logistic formula which then takes the probability of the result and specifying whether the result is 0 or 1. Random Forest is a voting machine learning model that uses the voting system between multiple decision trees in order to find a good prediction class (hence the name forest).

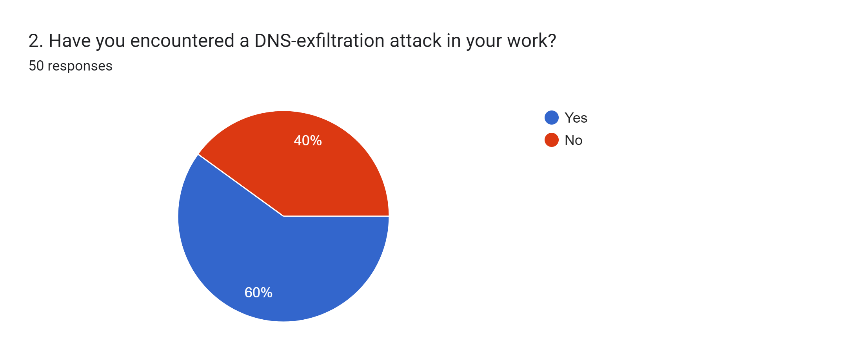
While the neural networks work by applying brain simulations on how networks connect by neurons and weights for each feature. I used RNN (Recurrent neural network), and it works by going through the inputs sequentially, through their recurrent connections in the networks. It is suitable for time-related problems, while LSTM (Long-Short term memory) is an updated and advanced version of RNN that helps grant the fix to the vanishing gradient problem that RNN suffers from. GRU (Gated recurrent unit) is also a way to handle the vanishing gradient problem and it is computationally more efficient than LSTM and RNN.

1. **Evaluation of models using metrics**: Since the problem is a classification problem, the common classification metrics will be used here, accuracy, precision, recall, and f1-score. This step is very crucial to the models because it decides whether the models are predicting the attack continuously with good accuracy or not. Not doing so can result in a bad model that is not recommended and a failed modeling process. But the accuracy gained was very high (95%) which can be considered very strong for a security problem, therefore recommended for the organizations to use and prevent DNS-Exfiltration attacks. As for the neural networks, they will be measured by the accuracy and the loss due to the need for a testing dataset to apply other measures.

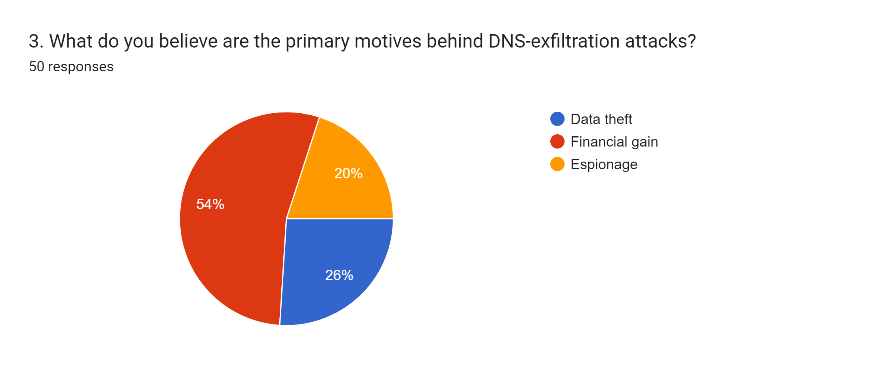
# Results and Discussion

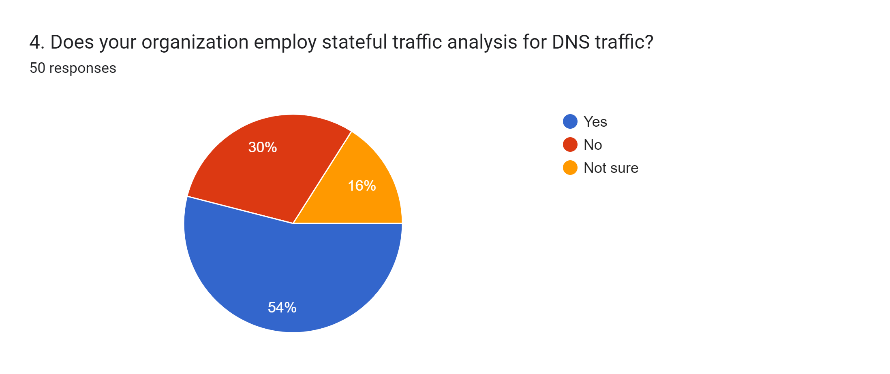
After applying two strategies to gather data about the DNS-Exfiltration attacks, I used plot and figures to gather the results of each method, and to communicate the researching progress through plots, because it can be clearer and more understandable this way. Each survey question had a plot to analyze the responses of the 50 experts. As the following:

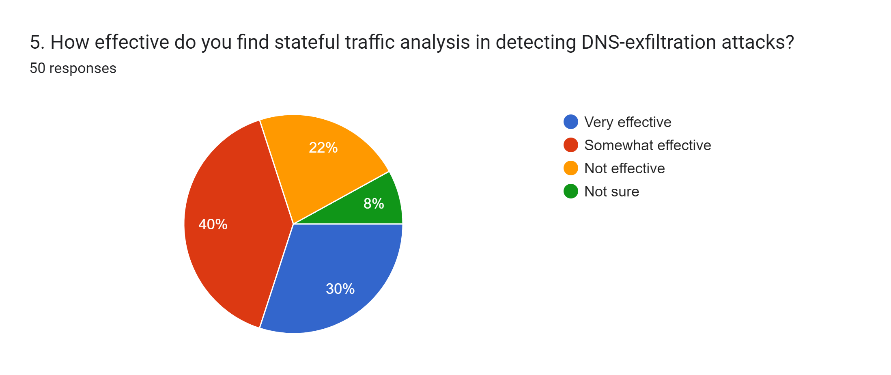
This question was aimed to see if the experts were familiar of such an attack that happens on the DNS traffic, the responses were neutral where nearly half the participants heard of it, a quarter of them were familiar with it, and the last quarter never heard of the attack.

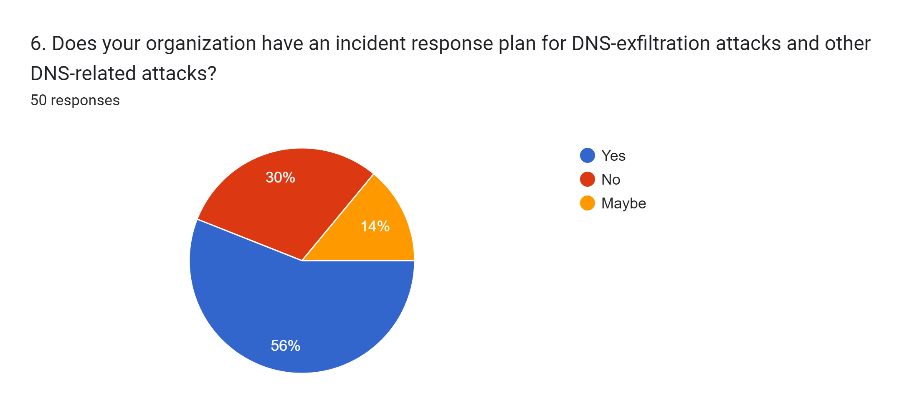


This question was regarding whether the organization has encountered a DNS-Exfiltration attack before, and the responses were also neutral where more than half of the experts happened to answer with Yes, and 40% answered with No.

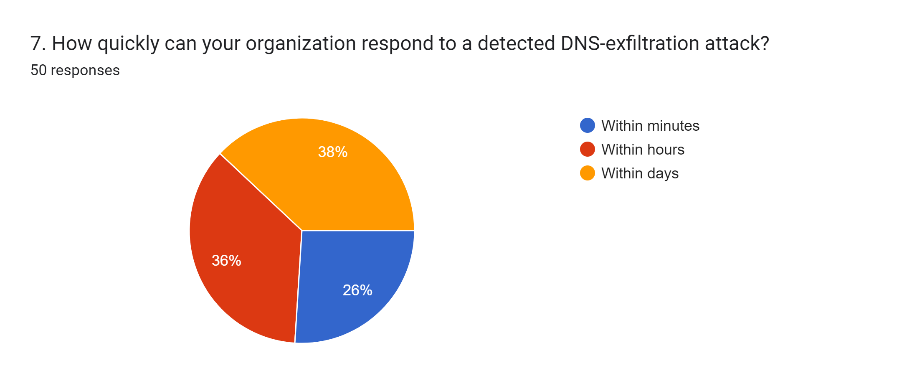
This question was pointed on what the aim of such attacks was in the thought of the experts, Most of the responses pointed that such attacks were done for financial gain, while the rest of the participants thought it was due to data theft reasons or espionage.

This question asked whether the organization uses the stateful traffic analysis to counter the attacks or not, most of the participants pointed that they use stateful traffic. While the rest weren’t sure about the usage.

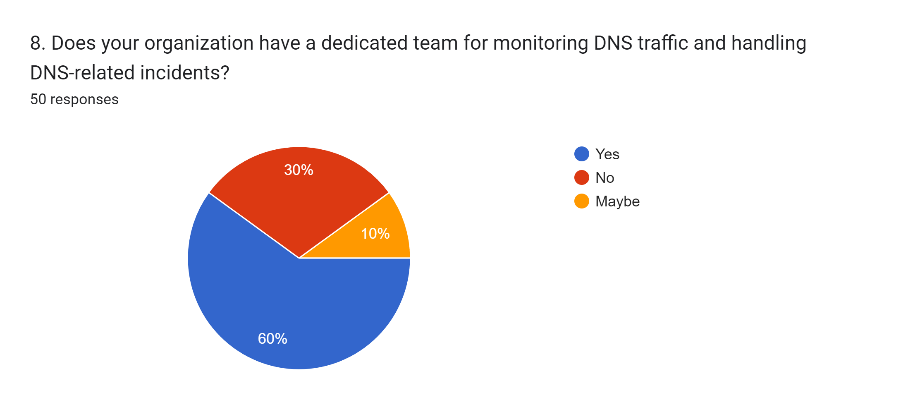
This question stated that nearly half of the experts find stateful DNS traffic analysis somewhat impactful in countering the DNS exfiltration attacks while 30% think it is very effective, the rest find the stateful traffic not effective at all.

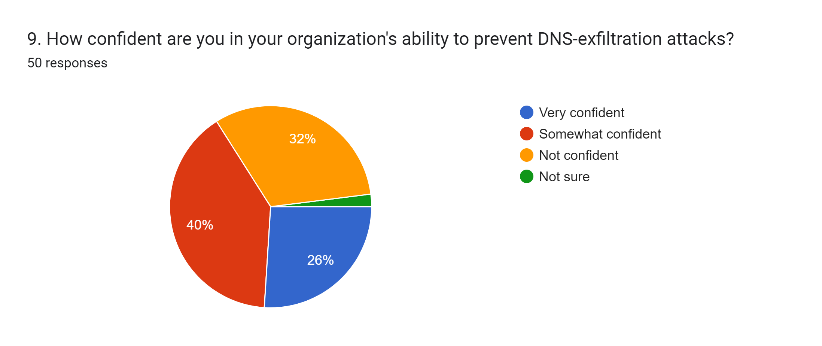


This question showed that half the experts believe that the institution has an incident response plan for such attacks. While the other half believes the organization isn’t ready to face such attacks.

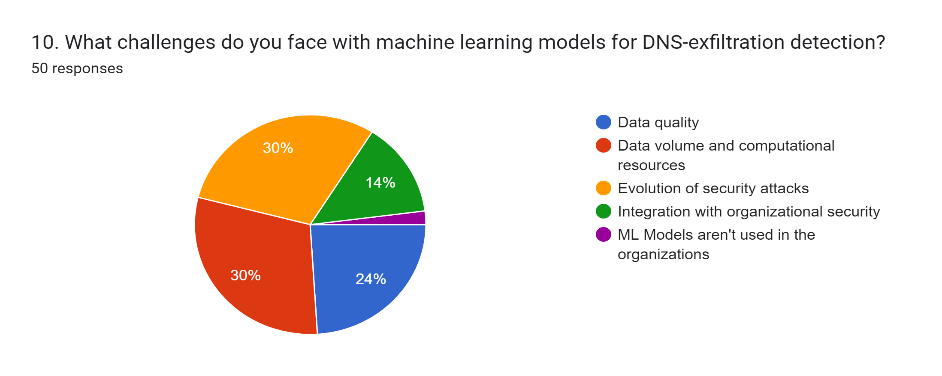


This question shows the response time in the beliefs of the experts inside the organization for the attacks. It is divided into three parts, the first part believes they can counter the attack within minutes, the second part within hours, and the last part believes that the attacks last days before getting countered.

This question asks if there is a monitoring team for the DNS traffic that handles the attacks, and the experts believe so, but a minority of them believe that the organization doesn’t have a team for monitoring the traffic.



This question shows how confident the professionals are with their organization’s ability to stop such attacks and counter them. The majority was somewhat confident, while 26% was significantly confident, and 32% wasn’t confident at all.

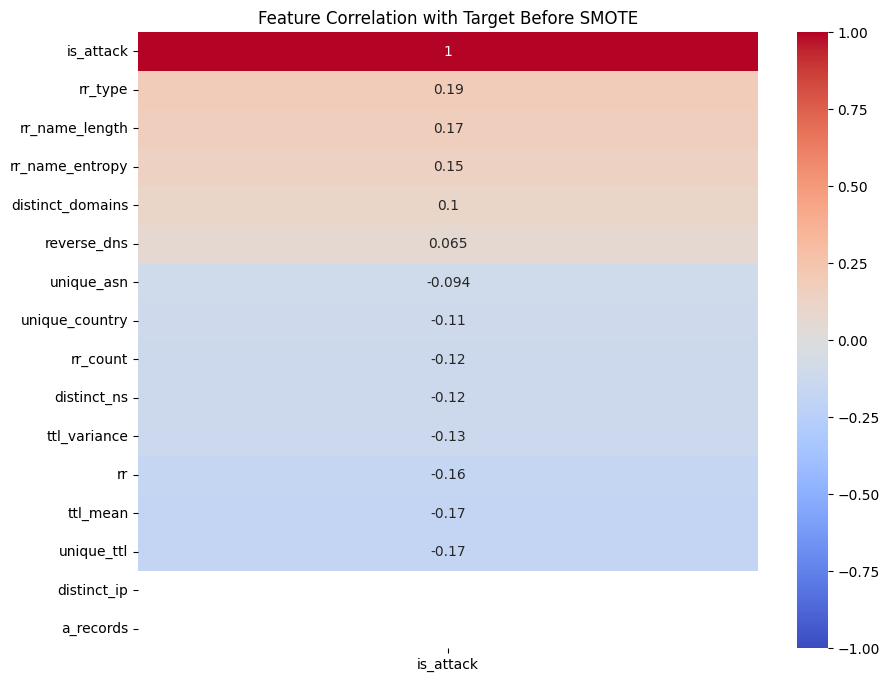


This question presumes the knowledge of the experts and asks them what would the problem be if ML was integrated into the detection of the attack. 30% believe it is due to the data volume and the computational cost of such implementation, while 30% of the individuals believe its due to the security attacks that may evolve beyond what machine learning algorithms can predict and prevent. 24% believe its due to the data quality and that the sensor data cannot be processed effectively. While 14% believe it is due to the organizational security itself being incompatible with the implementation of ML models inside the organization.

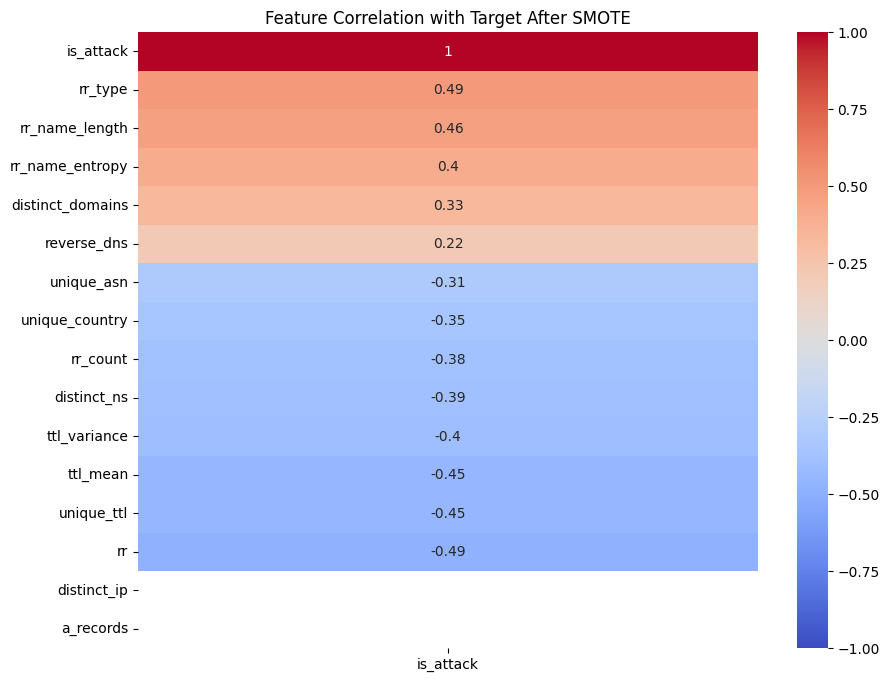
These were the analysis of the survey results according to the responses of the 50 experts in the institution. Now we will check the modeling evaluation metrics and discuss the results of the machine learning models and their predictions of whether there was an attack or not on the stateful traffic.

As for the python code, there were various results along its execution and completion, as following:

1. Heatmap results: The heatmap was used to study the correlation between the features and the target variable before and after the balancing of SMOTE, it was used using the Matplotlib library to provide more in-depth feature analysis in order to select the features relevant to the model and eliminate any features that aren’t correlated with the target varibale:



This heatmap was done before the balancing process, it shows that most variables are correlated poorly. We will now do the heatmap again after the balancing process:



This is the heatmap for the correlation of the features after the balancing process, it shows a stronger relationship between the features and the target variable which indicates that balancing is necessary to increase the correlation between features.

After doing the heatmap, and applying multiple preprocessing technique, I split the data into training and testing data, and started the modeling process for Machine learning models and Neural network models. The purpose was to evaluate and predict the attacks on the stateful traffic. So I will use four metrics that help give KPI for the models, which are accuracy, precision, recall and f1-score.

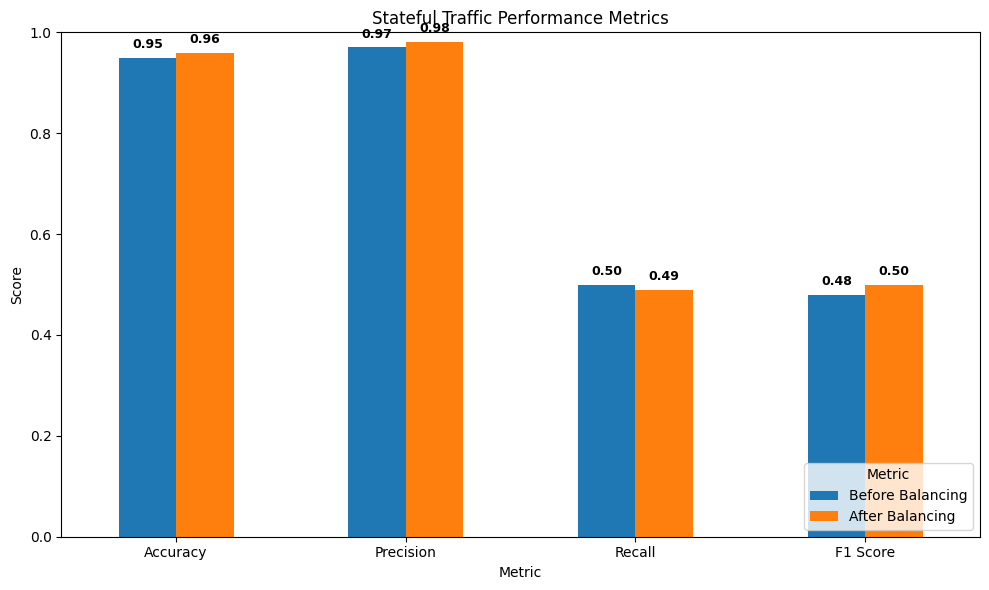
These were the average results of the machine learning models before and after balancing. This bar plot provides an overview of the stateful traffic’s analysis and its prediction of the DNS-exfiltration attacks. It shows that most metrics didn’t drastically change before and after balancing, and they were as the following:

- **Accuracy** of the models was 95% before balancing and 96% after balancing, which is considered significantly strong for detecting such attacks. It shows that the models can be used inside organizations to counter the DNS-Exfiltration attacks.

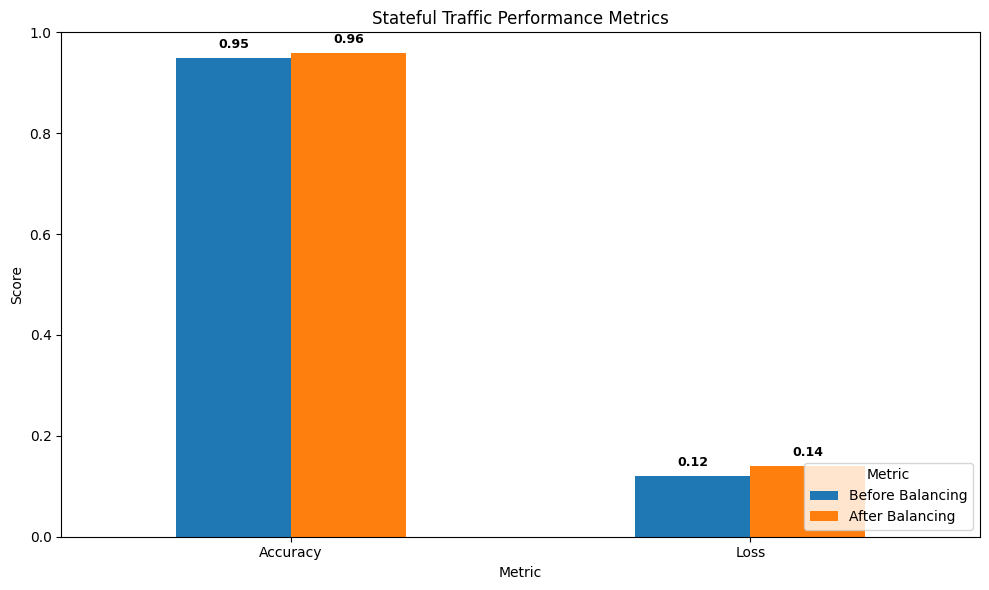
- **Precision** of the models was 97% before balancing and 98% after balancing, which shows that the models can mostly predict positive values, this is considered strong because the models can be reliable to predict the right class of the attack target.

- **Recall** of the models was 50% before balancing and 49% after balancing, this measure is considered weak, and it means that the models can’t predict the (1) values always. This means we can take the precision instead of the recall because of how low it is compared to it.

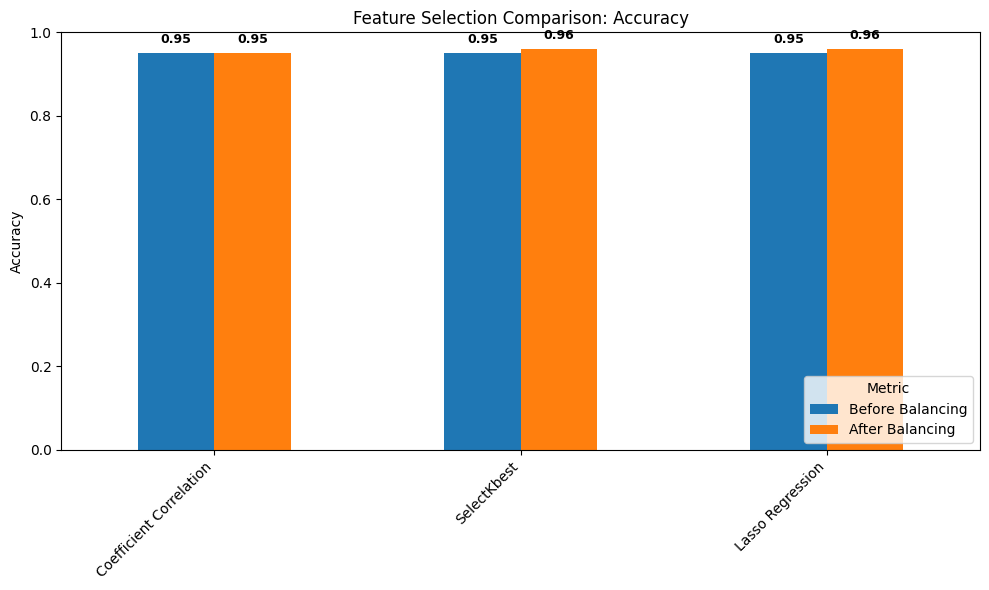
- **F1-Score** of the models was scored 48% before balancing and 50% after balancing, this is due to the harmonic mean between the precision and the recall. This score is considered weak but it is logical because of how low the recall score was. So we will take the accuracy measure instead of the f1-score.

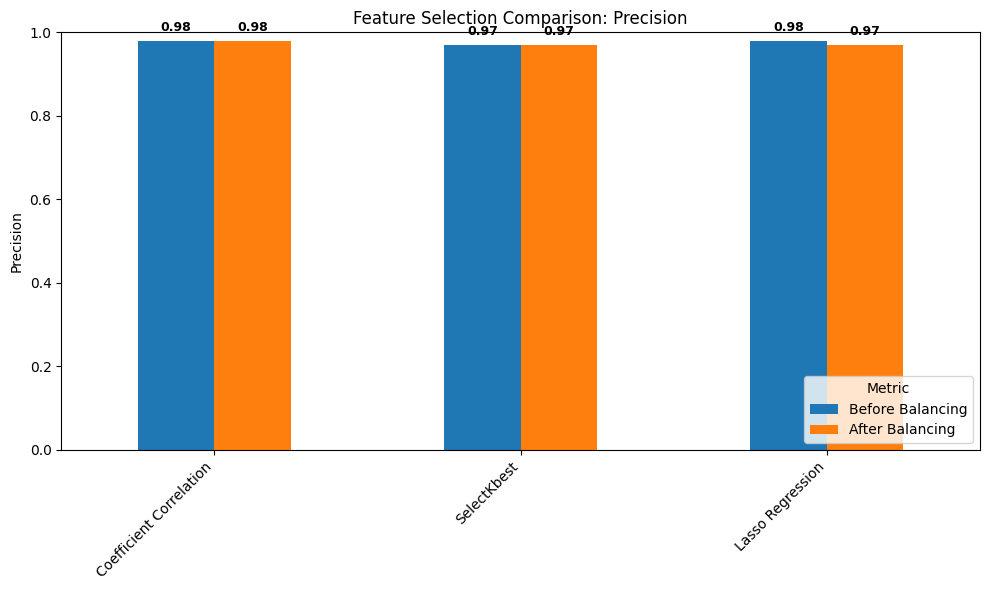


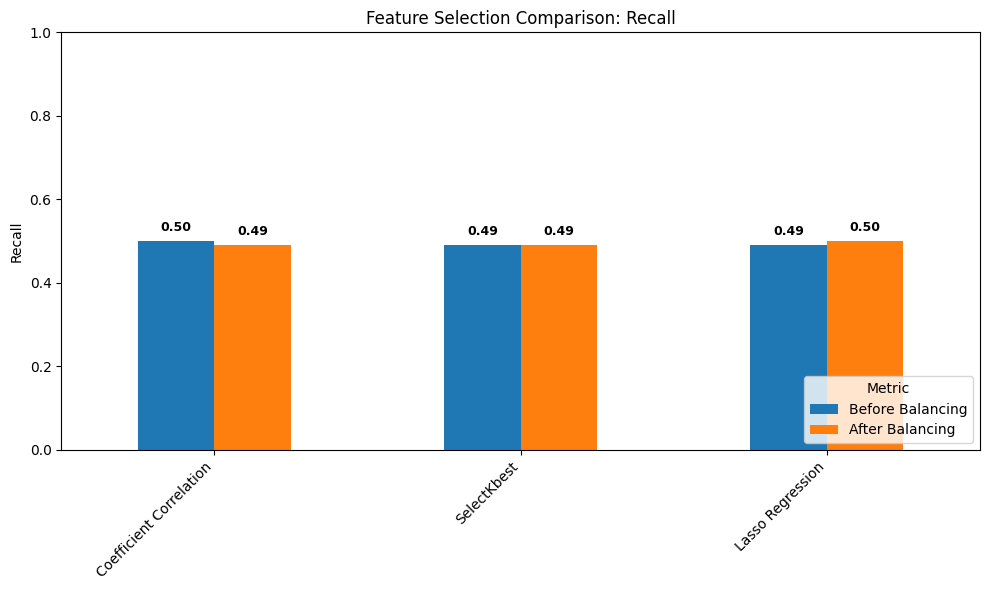
For the results of the neural network models though, they will be evaluated using two metrics only, accuracy and loss. Since there is no testing dataset for the DNS-Exfiltration. I performed manual selection of features with neural models (GRU, LSTM, and RNN) I took the average of the metrics and combined them in a bar plot:

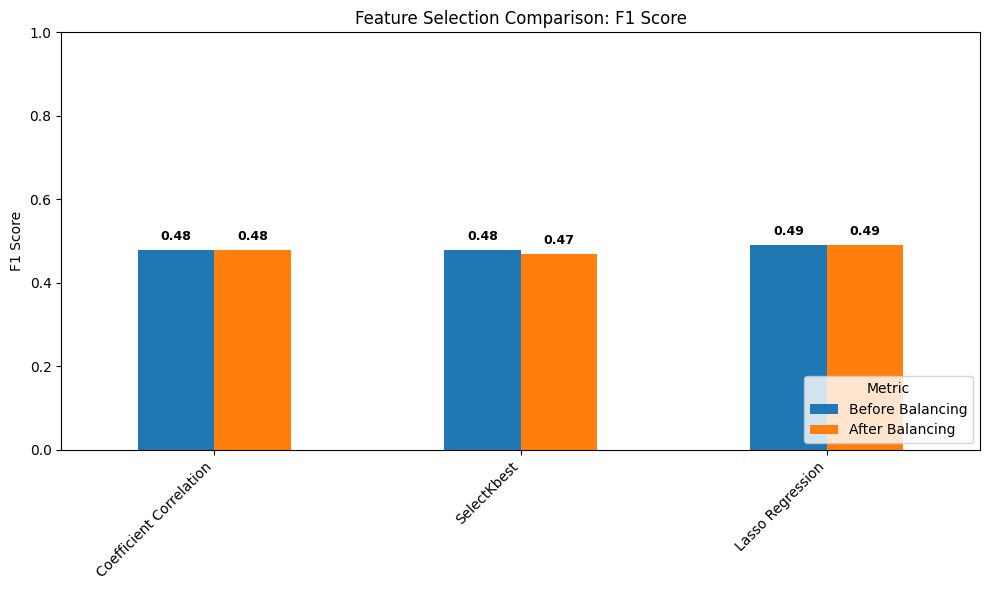


Then, I performed an automatic feature selection on the features, and these were the results (one plot for each metric).









Explain how the results meet the research question and objectives.

The results of each method were compared to both the research question and the objective, if both are achieved, then I can call my research progress a success. Starting with the research question:

**Based on the preprocessing and analysis of the features of the DNS-Exfiltration dataset. Can the deployed ML models predict the malicious attacks out of the benign traffic? While deciding whether the attacks can happen or not?**

This question studies whether the machine learning models can predict the attack occurrence or no, we can achieve this, by applying statistical measures on the models, and therefore generating an accuracy score. The accuracy score average was 95%. This means that the models are highly capable of achieving a good prediction on whether the malicious attacks would happen out of the benign traffic based on the features. The research question was achieved efficiently and effectively

Moving on to the research objectives, which are the steps that can guide me towards a valid research topic regarding the DNS-exfiltration attacks. Following the objectives:

- **Finding a data set that studies DNS Exfiltration attacks**. The dataset regarding the DNS-exfiltration attacks was found and collected inside an institution by sensor readings of stateful DNS traffic. It contained nearly 28 features and approximately 250,000 records. The goal was to predict whether an attack would occur or not.

- **Performing preprocessing steps and feature engineering methods for the data set, to involve only the most important features into the modeling phase**. Many preprocessing and feature engineering stages were implemented in order to produce a cleaned dataset empty of NULL values and with good feature selection methods. The features selected from feature selection techniques will then be added to the modeling stage.

- **Starting the modeling phase and studying the performance of the models used and comparing them to find the best model to use for future uses**. This objective was successfully achieved when I implemented the machine learning models on the clean dataset, to achieve good performance and accurate insights of the problem. It involved evaluating the models on four measures that were compared to in the end to find the best model and generate the average of the accuracy and the other metrics.

**- Conducting a survey (Primary data) for the research problem, to get feedback and notices on how it is explained and understood from the public, and its danger on the organizations, and plotting the survey results.** A survey was conducted on 50 participants which are cyber security experts in a specific institution. The survey had 10 close-ended questions that each was analyzed to dig further into the research problem and to validate it even more. It was shown that most experts know what the attack is about, but aren’t ready to face it or do not have the capabilities of countering such attacks.

Describe merits and limits of the analysis.

The survey analysis was done on a small number of experts within a single institution. This can mislead to false and limited information regarding the research topic. But the answers were conducted in an efficient manner. And the analysis was clear and concise, and the research was validated even further. The number of questions was short. Which means not every point was studied and further analyzed which can be time and effort consuming. But the archival research has helped me gather more analysis and further explain the DNS-exfiltration attacks and how they manage to counter such attacks using machine learning models and other ways.

The python code had a few merits and limits. First, the study was done only on stateful traffic analysis, which can be effective to prevent and detect the malicious DNS attacks, but working on stateless DNS traffic can be significantly effective and would further help prevent such attacks from occurring on the traffic. The code itself had a few ups and downs. Such as the data not being balanced and having to use a balancing technique to fix it. And some features were reduced and eliminated because they had no purpose and correlation with the target at all. The recall metric wasn’t great but we can rely on precision and other metrics to validate the models.

# Conclusion and Recommendations

State the aim of the study and summarize the findings.

The aim of the study was to find ways to detect and prevent the new form of DNS attacks that is attacking enterprises, the DNS-exfiltration attack. Since it acts as a peaceful client rather than an attacker, it generates a huge threat and targets the organizational data, and the financial state of the organizations. It must be detected before it attacks inside the DNS traffic, so that it can be prevented of damaging the enterprises. After using a form of survey and a python code experiment, I noticed some things:

1. The attack is still new to the experts in cyber security, which needs to be fixed quickly by implementing methods of awareness for this type of attack and spreading the word inside organizations too, because most organizations don’t have the ability to counter such attacks yet.
2. Machine learning algorithms with high-level preprocessing can detect the attack over a stateful traffic, by applying the models, organizations have a 95% chance of countering the attack before it starts damaging devices and steals information and money.
3. The response plan of some organizations is beyond dangerous, some organizations take minutes to prevent the DNS-exfiltration attack from happening, while some take hours or even days. This is due to the ignorance and evolution of such attack; most organizations haven’t heard about the attack until they are targeted by it.

Since the attack is still evolving, many organizations have yet to hear about its danger, and our goal is to spread awareness about this type of attack, and to provide guidance in order to detect it before it spreads onto the enterprise and destroys it completely.

Based on the findings, discuss the recommendations that you have.

1. It is recommended to conduct various seminars and awareness advertisements on this attack for the organizations that are being targeted by the attack, to prevent it from spreading and to further evolve, because most organizations have no idea what the attack is about or how dangerous it can be. It is a form of DNS attack that can lead to financial loss and data theft. It is serious and dangerous and should be known for all cybersecurity experts.
2. Using machine learning models is noticeably significant and necessary, due to the fact that the accuracy for detecting the attacks on stateful traffic has been immense with a percentage of 95% for detecting whether the traffic is an attack or not. This can help organizations detect and stop the attack from entering the devices, therefore damage the enterprise. It may need high computational power and an AI team to monitor such models, but protection for the organizations is necessary and essential.
3. Based on the findings, I have concluded three scenarios.
4. The machine learning models provided good metrics with the manual selection of features before balancing and after it. No automated feature selections or Neural networks are used here, this makes it time-consuming to manually select the appropriate features, and can lead to bias in choosing the right features, but it is computationally lower than the neural networks.
5. Using neural networks provided great metrics too, accuracy and loss were used in the evaluation part. It was also used with manual feature selection before balancing and after it. It can be time-consuming and computationally high to implement.
6. I used automated feature selections with machine learning algorithms, which produced good readings for the feature selections. This scenario revolves around automated feature selection process, which is time efficient, while being computationally low.

Based on these scenarios, I will choose scenario number 3, because implementing performance-supported machine learning methods, while having automated feature selection is very efficient for the organizations and for researchers to find data and conduct research on it. This is efficient and it helps gather data efficiently and quickly.

**Future work:**

I advise organizations to use the accurate models that are very efficient in detecting the DNS-exfiltration attacks via the stateful traffic sensors, since it can defend against such attacks and prevent the loss of data and financial damage for the facilities. I also advise them to further enhance the monitoring system because the experts’ opinions have been significantly negative on its performance to detect this type of attack. I would recommend organizations to conduct more study on the DNS-exfiltration attacks, and further allow experts to be knowledgeable on this cyber attack because it can negatively impact the state of the organization. Organizations should invest in purchasing a better technological system to counter such attacks, because most organizations disregard its danger and ignore it. It needs to be prioritized and never ignored. Starting with better powered devices and secure systems, and better databases to hold the data and more accurate sensors for traffic.

# Reflections

Avoid generalization and focus on personal development and the research journey in a critical and objective way.

### Selected Research Methodology

The research process has been a continuous learning trip for me, where I understood the importance of selecting the right methodologies to start and work and end the research with. I was aware of the limitations that come with the selected methodologies and I had more alternate plans in case the original research process structure doesn’t work. I used Realism and Positivism to understand the approach of the DNS-exfiltration attack research that I’m taking. I used an Abduction theory development approach so that I can generate predictions based on theories that I have read through journal papers about the DNS-exfiltration attacks. I used a mixed-method simple approach that combined both qualitative and quantitative approaches to collect data for the research, I used surveys and experiments to collect data from individuals and sensors regarding the DNS-exfiltration attacks, and I had them collected in a cross-sectional approach which happened efficiently.

1. **Realism & Positivism**: I combined both methodologies to understand the nature of the research I’m conducting in a more effective manner, where I focused on identifying the general laws of the research, and acknowledging that it exists in a reality whether its conflicting with people’s knowledge. But it had limits where sometimes individuals’ opinions would be disregarded or ignored completely, which can be bad because they might have more information about the research I’m doing (DNS-exfiltration attack)
2. **Abduction**: this methodology allowed the research to move from theories and hypotheses to actual real predictions and information for the DNS-exfiltration attacks with evidence. It was appropriate for my use of machine learning algorithms within the research, but relying on predictions can be inaccurate for some research topics. Since it wouldn’t give a definitive result as would a deductive approach do. Which can result in lower credibility for the research itself.
3. M**ixed-method simple**: This approach was used to take the advantage of combining both qualitative and quantitative data collection methods for the research, since it would give a wider view for the research, and more information for the sake and validation of the research. It is however simple, and may not give complex and detailed information for the DNS-exfiltration attacks. There may also be some form of bias when combining both methods, as one may ignore the results of a method and focus on the other.
4. Usage of **surveys**, **experiments**, and **archival research** helped with enriching the research with information about DNS-exfiltration attacks, whether it was historical data on the attacks or insights from current cyber security experts in enterprises, or inference using data from traffic sensors. The goal was to combine the information from the three methods and validate the research using the results and evaluations of the methods. But a challenge was encountered when combining them, to avoid biased results and balancing the three methods together to achieve a suitable and effective contribution towards the research of the attack. It took some serious consideration and detailed planning to counter this challenge.
5. **Cross-sectional time**: The use of this methodology was seen from how data collected was made, surveys and experiment results were collected in a single time-frame which gave a stable and efficient research state. Despite the efficiency however, this methodology is unable to keep track of time changing, which would make it a challenge to be observant of trendy attacks and evolution of the DNS-exfiltration attack itself. Which would make it unable to recommend to facilities.

### Alternative Research Methodologies

The research methodologies selected were efficient to produce valid research on the future of DNS-exfiltration attacks and ways to prevent and detect such attacks using stateful traffic analysis. But more methodologies can be used to enhance and change the research process for future research conducted on the topic. Which can greatly expand the knowledge and personal understanding of the research dynamic and how researchers approach new ideas with new ways daily. It can be an open door for new information and knowledge. These are some of the alternate methodologies that I may use for future research processes:

1. **Action research**: This methodology is used when you would want cooperation with the stakeholders in your research. It would lead to more efficient results and would bypass any security regarding the DNS-exfiltration, which would lead to more in-depth analysis and wider overview for the problem and the research conducted. This cooperation would be a platform for sharing practical and theoretical knowledge on the DNS-exfiltration attacks and how to tackle and counter them.
2. **Longitudinal study**: This methodology uses continuous periods of time to observe changes over time. It can be used to provide a deeper analysis and understanding of the DNS-exfiltration attacks and how they’re prevented. This can be very effective in the experiment data collection method since it can observe and analyze the evolution of the threat and would help gather information on new patterns in the stateful traffic and the attack itself.
3. **Mixed multi-complex**: The mixed multi approach was used in order to find and combine the best qualitative and quantitative approaches for data collection methods. Use of the complex variation can lead to diverse sources to collect data from, and would offer more comprehensive insights that would lead to further validation for the research. This can enable more analysis, and blend more methods together to reach a fully efficient conclusion with detailed information on how to deal with DNS-exfiltration attacks on a new and improved stage.
4. **Ethnography**: This study can help gather more information and observe the daily job of the experts in the cyber security departments that counter the DNS-exfiltration attacks, to find how they’re detecting and countering them, and then provide more insights and information on how to help further eliminate such threats. This method can lead to interactions with the experts themselves which can influence different cybersecurity practices.

Many lessons were learned throughout the research conducted, and the selection of different methodologies using the onion model, which have improved my skill in conducting research efficiently with lower issues and increased performance for future researches. Such as:

1. Noticing the limitations while selecting the methodologies is always necessary, so that you would expect the outcome of such negative results. The adaptability to methodologies while having knowledge of their downside is the key to a successful research process and increased continuous adaptation in the nature of the research field, in my case cyber security and artificial intelligence.
2. The integration between data sources (Quantitative and qualitative) was necessary and important for the research, since finding new ways to gather information and data regarding the DNS-exfiltration will benefit the research and validate it even further and make it more robust for more researches.
3. Modern technology is needed to research the DNS-exfiltration attacks and prevent them from damaging the organizations. Such technology can be higher computational powered devices, servers for data, more advanced and accurate sensors, more readable structured databases. Such attacks can be shortened and countered more often with the right technology

### Recommended Actions and Future Considerations

After conducting the research, and understanding the nature of the DNS-Exfiltration attacks, it came to mind some considerations that I would work with and recommend other cyber analysts and experts to follow too, especially big enterprises and institutions. Such as:

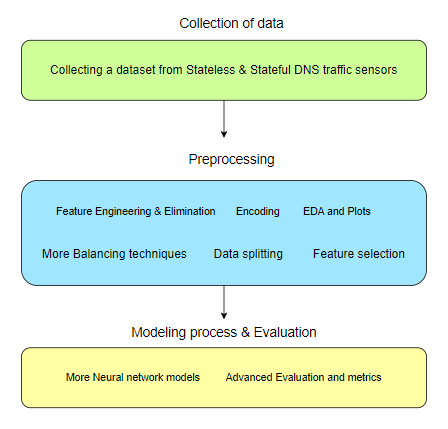
1. Investing in technology, because continuous exploration for new, dangerous, and advanced DNS attacks are frequent, and organizations need to be aware and ready to prevent such disasters from happening. Some examples on technologies are:
   * Advanced servers
   * High computational powered devices and PCs.
   * More structured databases to hold the data.
   * Better monitoring technology and team.
   * Advanced sensors for the traffic.
2. More awareness for cyber security enthusiasts, since the survey conducted has shown that most experts inside facilities aren’t aware of the danger that’s targeting organizations, which is the DNS-exfiltration attacks. This can be done by holding seminars and more awareness campaigns that can benefit and allow experts to be knowledgeable in the field.

Here are some recommendations that I would recommend for researchers, and actions that I would consider doing in future researches, to deeply evaluate my researching skill and to increase my research validation:

1. Selecting flexible and more suitable methodologies when conducting future researches, because the research would have the ability to adapt to new and emerging challenges that would endanger the validity of the research. This would open the door for new information that would further enhance the nature of the research.
2. Conducting more frequent surveys with a wider audience, and interviews, since it can lead to new opinions and more information regarding the research topic. Which can strengthen the research conducted, and encourage new contributions to the research.
3. Within the DNS-exfiltration attack python code, I would recommend using more ways to evaluate and predict whether an attack would happen or not, I chose machine learning algorithms for this task. Using neural network models would benefit the research to produce more accurate predictions for bigger datasets.

### Recommended Methodology

These are the methodologies that I would recommend to new researching contributions to the DNS-exfiltration research with stateful traffic analysis.



1. **The collection of the dataset through sensors would need to hold both** **stateful and stateless traffic**, to further understand the DNS-Exfiltration attack even more, by studying the stateless traffic, the research would then be conducted on both DNS traffic types. Which will improve the state of the research.
2. **Feature engineering and elimination**, since we have a bigger dataset now, it would be necessary to reduce the irrelevant features via elimination. By dropping them from the dataset, to produce new and improve results that would benefit the research.
3. **Encoding.** Encoding transforms the categorical features (Objects and Strings) to Numerical features, so that they can be trained effectively into the models, I removed the scaling part, because in this methodology, we will be using neural network models, which produce a scaling process on their own without the need to scale them manually.
4. **EDA & Plots,** this stage is necessary to figure out the state of the dataset for both stateful and stateless traffic while focusing on the correlation between the target variable and the independent features themselves, in order to verify whether the features are relevant or not.
5. **Different balancing techniques** may be used in order to maintain the balancing of the dataset, last time, SMOTE was used as a balancing technique to balance the minority class from the target variable with the majority class. To further enhance the research process, we may use different balancing techniques that may balance the datasets in different ways and implementations. So that more information would be generated in an efficient manner, such as random oversampling and ADASYN.
6. **Data splitting (Training & Testing),** after balancing and preprocessing the data, it is time to prepare the data by splitting them into training and testing data, so that they can be trained and fit into the neural networks and compared then with the testing data. To produce results and accurate insights.
7. **Feature selection** is also used in this methodology to counter any irrelevant and useless features that are not needed in the training process, since such features may decrease the performance of the neural networks, and produce false accuracy.
8. **Neural network modeling** is one of the most common modeling processes for bigger datasets, it supports data volume, variety, and velocity. The need for more computational power is necessary for such modeling processes. Since it is trained on neural networks such as the human brain. It mimics the brain to predict various tasks based on the same neuron systems inside the brain. DNS-exfiltration attacks would be detected and prevented in a more efficient way using neural network modeling. Such as MLP, Bert, GPT, CNN and more.
9. **Evaluation of models using more advanced metrics**, such as the common four metrics, accuracy, precision, recall and f1-score, with the addition of log loss, specificity, AUC and ROC. And other advanced metrics that can help produce a better understanding of the modeling process for the DNS-exfiltration attacks under the improved system.

# References:

[1] A. Melnikovas, “Adapting Research Onion Model for Futures Studies,” [https: //www.semanticscholar.org/paper/Towards-an-Explicit-Research-Methodology-% 3A-Adapting-Melnikovas/dccf54a5a4312ceb2261e1989cda01f73989d735,](https://www.semanticscholar.org/paper/Towards-an-Explicit-Research-Methodology-%3A-Adapting-Melnikovas/dccf54a5a4312ceb2261e1989cda01f73989d735) 2018, [Online; accessed 10-Jan-2024].

[2] Samaneh Mahdavifar, Amgad Hanafy Salem, Princy Victor, Amir H. Razavi, Miguel Garzon, Natasha Hellberg, and Arash Habibi Lashkari. 2022. Lightweight Hybrid Detection of Data Exfiltration using DNS based on Machine Learning. In Proceedings of the 2021 11th International Conference on Communication and Network Security (ICCNS '21). Association for Computing Machinery, New York, NY, USA, 80–86. <https://doi.org/10.1145/3507509.3507520>

[3] J. Diao, B. Fang, X. Cui, Z. Wang, T. Wang and S. Song, "From Passive to Active: Near-optimal DNS-based Data Exfiltration Defense Method Based on Sticky Mechanism," 2022 IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), Wuhan, China, 2022, pp. 159-166, doi: 10.1109/TrustCom56396.2022.00032.

[4] Shashank Biradar, Shramik S Shetty, Pradeep Nayak, Prajakta Shetty and Shwetha R Sharma (2022). A Detection Method Against DNS Cache Poisoning Attacks using Machine Learning Techniques. *International Journal of Advanced Research in Science, Communication and Technology*, pp.671–675. doi: https://doi.org/10.48175/ijarsct-7033.

[5] Saha, S., Moinul Islam Sayed and Islam, R. (2023). Detecting DNS over HTTPS Traffic Using Ensemble Feature-based Machine Learning. *Chapman and Hall/CRC eBooks*, pp.118–131. doi: https://doi.org/10.1201/9781003256083-10.

[6] Izabela Savić, Yan, H., Lin, X. and Gillis, D. (2024). Adversarial Example Attacks and Defenses in DNS Data Exfiltration. *Communications in computer and information science*, pp.147–163. doi: https://doi.org/10.1007/978-981-99-9614-8\_10.

[7] Kumar, H. and Das, H. (2023). Software Fault Prediction using Wrapper based Feature Selection Approach employing Genetic Algorithm. doi: https://doi.org/10.1109/otcon56053.2023.10113911.

[8] Sumi and Narayanan, A. (2019). Improving Classification Accuracy Using Combined Filter+Wrapper Feature Selection Technique. doi: https://doi.org/10.1109/icecct.2019.8869518.

[9] Gnana, A., A. Escalin Fernando and E. Jebamalar Leavline (2016). Experimental study on feature selection methods for software fault detection. doi: https://doi.org/10.1109/iccpct.2016.7530156.