Sheffield Hallam University | College of Business, Technology and Engineering

MSc Dissertation Report

DETECTION OF MALICIOUS URLS AND FINDING THEIR ATTACK TYPES USING AI TECHNIQUES

A dissertation submitted in partial fulfilment of the requirements of Sheffield Hallam University for the degree of Master of Science in Artificial Intelligence

Student Name	Ashika Srinivasan
Student ID	32066221
Supervisor	Ms. Caren Fernandes
Date of Submission	22/01/2024

STATEMENT:

This dissertation **IS CONFIDENTIAL** and circulation should be restricted to those involved in its assessment only.

Acknowledgements

I would want to dedicate my deepest gratitude to my dissertation advisor, Ms. Caren Fernandes, whose advice and strong support were important throughout the dissertation research process. Also, for your insightful feedback and helpful suggestions, which greatly improved the quality of the work. Special thanks to Mr. Jing Wang, a potential beneficiary who generously volunteered his time and knowledge to provide valuable comments on this dissertation project. Finally, I am grateful to Sheffield Hallam University for providing the essential resources and environment for research.

Abstract

Economic damage, reputational harm, and even national security breaches might result from cyber assaults that originate from rogue URLs. The purpose of this project was to create an effective system for detecting faked URLs and recognising cyber threats by combining cybersecurity concepts with new techniques for analysing data to improve threat detection accuracy and contribute to better cybersecurity measures. The inductive approach that combined qualitative and quantitative research techniques allowed this study to achieve its aim. Secondary sources were utilised to gather data for the research. In this regard, a list of URLs classified as phishing or malware is compiled from multiple web sources as part of the data-collecting process. Data collection also makes use of the Kaggle website. The findings of the research show that, when compared to traditional methods, DL and NLP algorithms are much better at detecting bogus URLs. URL-based phishing, email spam recognition, virus identification, and website defacement detection tasks have all been outperformed by DL algorithms such as DNN and CNN. However, some obstacles must be overcome, such as choosing the optimal DL method, keeping the training database current, and gathering enough training data. The research goes on to say that characteristics including safe HTTP protocols, URL size, character assessment, odd subdivisions, and unusual subdomain assessment are crucial for identifying fraudulent URLs. Outcomes from comparing four different classification methods (ANN, Random Forest Classifier, Logistic Regression, and Gradient Boost) shed light on their relative merits and point the way toward potential new avenues of investigation in the field of cybersecurity. To enhance cybersecurity applications' capacity to identify and address risks presented by harmful URLs, it may be beneficial to combine DL and NLP approaches with data pretreatment and exploratory analysis of data.

Table of Contents

Chapter 1: Introduction

1.1 Introduction

The widespread growth of the internet has had a significant effect on transforming communication and commercial operations for both individuals and organisations. However, it is also coupled with the increasing risk of cyberattacks, particularly through hazardous Uniform Resource Locators (URLs). Cybercriminals use these URLs as entry points to initiate attacks, compromise internet security, and exploit vulnerabilities, posing a danger to online trustworthiness and security. Cyber threats arising from malicious URLs have the possibility to cause considerable financial loss, harm to reputation, and even compromise national security. Hence, it has become critical to detect rogue URLs and understand the exact assaults they enable to preserve the trustworthiness and security of the digital environment.

This study investigates, creates, and applies efficient techniques and tools for identifying malicious URLs and classifying the particular attacks they carry out in response to these issues. The purpose is to improve internet security by protecting users from potentially harmful websites. This study is motivated by the need to help keep ahead of the constantly shifting cyber threat landscape. In addition to helping people and businesses protect their digital possessions, this study also advances the subject of cyber security as a whole. It takes the use of modern tools like machine learning, and data analysis to deliver a thorough and up-to-date answer to an urgent problem, building on previous research. The chapters in this dissertation delve into the research methodologies, findings, and recommendations for improving malicious URL detection and improving understanding of attack types to strengthen cyber security.

1.2 Aim and Objectives

Aim

The major goal is to generate an effective system capable of detecting forged URLs and classifying attacks by integrating cyber security concepts with data analysis for enhanced threat identification accuracy.

Objectives

- To appraise the literature on malicious URL detection and attack categorisation approaches.
- To construct and build a powerful system capable of detecting harmful URLs using various methodologies and classifying them based on attack types such as malware dissemination and phishing.
- To evaluate the proposed system's real-world effectiveness and performance over malicious URLs.

1.3 Research Question

1. How can artificial intelligence (AI) technologies be effectively applied to detect malicious Uniform Resource Locators (URLs), while also differentiating and classifying the various types of cyber-attacks they might facilitate?

1.4 Background

The Internet has become an essential aspect of modern life, facilitating everything from instantaneous global communication and data storage to online shopping and financial transactions. Unfortunately, the rise of harmful URLs and the sophisticated assaults they facilitate is an unintended consequence of this digital revolution. Since its creation, the internet has gone a long way, expanding from a means of scholarly and military communication to a worldwide network that influences contemporary life in various areas. As the internet has evolved, so have the strategies used by cybercriminals. The first phishing attempts appeared in the early days of the World Wide Web, marking the beginning of malicious URLs. As access to the internet increased to include more people, companies, and gadgets, so did the potential for abuse. Phishing, virus distribution, identity theft, and data breaches are just some of the cyber assaults that have used malicious URLs as a delivery mechanism.

The widespread use of harmful URLs poses serious risks to both people and businesses. Cybercriminals are becoming increasingly sophisticated in their attempts to trick visitors into accessing malicious websites. As cited by Alanazi & Gumaei (2023), phishing attacks, for example, often use genuine URLs to lure victims into disclosing personal information like login credentials. However, malware dissemination takes advantage of people's trusting nature by tricking them into visiting harmful websites using safe links. The use of malicious URLs extends beyond the criminal community. Such URLs were used by state-sponsored attackers to compromise the government and infrastructure. There have been instances when these assaults have resulted in significant downtime, spying, and monetary losses. The magnitude of these dangers highlights the paramount need for solid, efficient defences.

The incidence of fraudulent URLs has increased, leading to the development of numerous tools for detecting and blocking them. As mentioned by Aljabri *et al.* (2022), signature-based and behaviour-based techniques are the two main buckets into which these strategies are placed. To identify and block harmful URLs, signature-based detection uses patterns or signatures that have already been identified. However, this approach is not perfect since it cannot identify "zero-day" assaults that have never been observed before. In addition, malicious actors often adjust their methods to fool signature-based defences. Instead of relying on static signatures, behaviour-based detection examines how URLs behave, including how they interact with people and what kinds of material they host. The advancement of behaviour-based detection systems has been greatly aided by machine learning techniques. As illustrated by Atif *et al.* (2023), because of their capacity to learn and adapt from massive datasets, these algorithms identify previously unseen harmful URLs by analysing their behaviour and attributes.The identification of harmful URLs also has major privacy implications. It is common practice to gather and examine the content of the web pages that a given URL leads to as part of the analysis process. However, there are certain issues with data security and user privacy related to such practices. In addition, threats are becoming more sophisticated at hiding their harmful URLs, making them harder to spot. They use methods like domain-generating algorithms, polymorphic malware, and URL-shortening services to avoid being caught by conventional security measures. To keep up with these strategies, machine learning algorithms must undergo constant improvement.

1.5 Problem Statement

The widespread use of malicious URLs has become a serious and pervasive danger to the safety and stability of the Internet as a whole. Many malicious cyber-attacks, such as phishing, virus distribution, identity theft, and data breaches, use these URLs as entry points.

A major difficulty in cyber security is developing reliable methods for detecting harmful URLs and, more importantly, for determining the exact attack types that they allow.

The sheer number and variety of harmful URLs that plague existing security solutions pose a major challenge. As cited by Coyac-Torres *et al.* (2023), cybercriminals are using innovative ways to hide malicious links in apparently innocuous domain names. Due to the constant evolution of these risks, many pre-existing security measures are incapable of mitigating the challenges posed by malicious URLs. As a result, there is an urgent need for the creation of more reliable and flexible detection strategies. In addition, there are severe repercussions for organisations that fail to identify bad URLs and the assaults they facilitate. People risk having their money, identity, or personal information stolen by cybercriminals. The stakes are considerably high for businesses, which are at a greater risk of suffering huge financial losses, reputational harm, and data breaches. When it comes to government agencies and other essential services, it threatens critical infrastructure as well as sensitive information, which can compromise national security.

Even though traditional detection methods such as blacklisting and signature-based approaches, which rely on creating unique signatures for previously identified malicious URLs, have proven useful in the past, these techniques are identified as struggling to detect new and unknown malicious URLs for which there are no unique signatures (Darwish *et al.*, 2023). Even the more adaptable detection technique, such as behaviour-based detection, has problems, including skewed data, a considerable number of false positives, and privacy concerns. A more robust and precise defence against rogue URLs cannot be developed without first overcoming these obstacles.

The study of URLs and their contents raises significant privacy problems. There are ethical and legal concerns surrounding user privacy and data security since many existing detecting technologies entail collecting and analysing online content. It is a complex problem that calls for creative solutions to find an effective medium between efficient detection and user privacy.

1.6 Rationale

Improving cyber security in the digital age is an absolute necessity, not a choice. In the present modern era, where the internet access has grown into an integral component of everyday life, fortifying digital defence has emerged as an important requirement. The need to fortify digital defence is at the heart of the motivation behind studies aimed at finding the harmful URLs and the identification of the attack types associated with them using machine learning algorithms. The justifications and primary motives for this study are discussed below.

First and foremost, the internet has become a breeding ground for a variety of cybercrimes due to its widespread usage in contemporary life. As illustrated by Dutta (2021), one of the most devious ways that hackers are more likely to breach the security of people, businesses, and critical infrastructure is via the use of malicious URLs. The widespread use of these URLs presents an immediate risk to the privacy and confidentiality of data stored digitally. In light of this, the study provides a preventative measure against bad URLs and the attacks they facilitate.

Apart from privacy concerns, this study's motivation is far more general. The study has the potential to make important contributions to the larger area of cyber security, given the global consequences of cyber-attacks. The outcomes and methods used in this study can have far-reaching consequences for shoring up digital defences. Collaboration in research and the development of novel countermeasures are significance by considering the global character of cyber threats, which impact a wide range of governments and organisations.

1.7 Scope of the Study

The parameters of this study specify its scope, which is the application of machine learning techniques to identify malicious URLs and classify the attacks that they are linked to. This defines the scope of the investigation and makes the research's aims more transparent. The main objective of this research is on enhancing existing methods and creating new ones for detecting harmful websites. This includes both conventional dangers and new ones that have never been observed before. This research goes beyond just cataloguing the prevalence of such dangerous URLs and examines machine learning algorithms to categorise the various attacks they enable. This larger lens helps shed light on the many forms that cyber threats might take. Furthermore, this study intends to extract useful tips and procedures for enhancing online safety and reducing the dangers posed by rogue URLs. These suggestions will be based on the study's results and conclusions.

1.8 Benefits of the Project

The project improves cyber security by correctly identifying malicious URLs and the various attack techniques they use. This preventative measure lessens the impact of attacks and reduces the likelihood of harm or data breaches. Understanding how attacks work allows for more effective responses, resulting in less downtime and expenses. By creating a safe online space for people and businesses, this initiative helps save money and earns confidence from its users by keeping their private data and assets safe.

1.9 Dissertation Structure

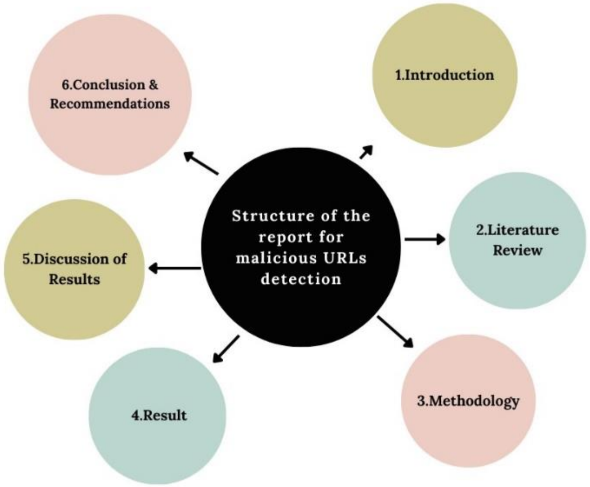

Structure of the report for malicious URLs detection
1.Introduction
2.Literature Review
3.Methodology
4.Result
5.Discussion of Results
6.Conclusion & Recommendations

Figure 1.8.1: Structure of the report for malicious URL detection

(Source: Self-Created)

Chapter 2: Literature Review

2.1 Introduction

The cyber security industry continues to deal with the issues posed by malicious URLs, which are sometimes hidden within seemingly safe web connections. These URLs act as entry points for various cyber-attacks, such as malware distribution and phishing campaigns. Understanding the methods employed by deceptive URLs is vital for protecting sensitive data, personal privacy, and the interconnected digital infrastructure.

This chapter extensively explores the current literature on detecting and categorizing malicious URLs. It sheds light on diverse methods and tools such as signature-based analysis; blacklisting; heuristic analysis; machine learning; behavioural analysis; and deep learning, developed to identify malevolent web addresses and counter the ever-evolving cyber security threats.

2.2 Malicious URL Detection Techniques

In this section, the various methods employed to identify and combat detrimental URLs have been explored, with particular consideration given to the characteristics that set them apart and the effectiveness of these defensive measures. These characteristics encompass recognised patterns, database matching, rule-based examination, adaptability through learning, and the automatic derivation of intricate data. As per Mohammed *et al.* (2022), to recognise harmful URLs, the signature-based examination is one of the most extensively used and time-honoured approaches. This approach involves contrasting a URL with a repository containing recognised patterns or signatures of malicious URLs.

A harmful URL is labelled as such when a correlation is identified. Antivirus software and intrusion detection systems (IDSs) depend greatly on this correlation-based identification approach since it is extremely effective in detecting typical cyber risks. Nevertheless, it is unable to cope with modern versions of perilous URLs or those that have never been encountered previously. In blacklisting, incoming URLs are matched against databases containing links to recognised hazardous websites. As cited by Nagy *et al.* (2023), a website's URL is deemed malicious if it is indistinguishable from the one on the blacklist. While blacklisting is a straightforward and efficient method to identify familiar dangers, it possesses certain constraints, like being unable to acknowledge novel, formerly unidentified detrimental URLs, and it might be challenging to maintain an updated blacklist.

Uçar et al. (2019) highlighted the advantages of convolutional neural networks (CNNs) and recurrent neural networks (RNNs), two major models in deep learning, in effectively detecting fraudulent URLs through a comprehensive study of both their layout and content. This suggests the importance of deep learning in mitigating the risks posed by deceptive URLs. Their work demonstrates a high level of accuracy in detecting fraudulent web pages. According to Prasad & Rao (2021), these algorithms automatically extract intricate data from URLs.

2.3 Machine Learning in Cybersecurity

2.3.1 Machine Learning Applications in Cybersecurity

Machine learning methods are identified as powerful tool for detecting irregularities within extensive datasets, such as anomalous patterns and outliers within a large collection of data. This ability is employed in the realm of cybersecurity to detect unusual patterns and behaviours that might indicate an imminent danger. Machine learning algorithms can offer a foundation of customary network conduct, empowering security teams to promptly respond to potential dangers by recognising and marking deviations from the standard. Intrusion detection systems (IDS) largely depend on machine learning to identify indicators of attacks or other harmful behaviour within a network, as illustrated by Rasheed et al. (2021). Machine learning models can detect malicious actions, even if they have never been seen before, by continually analysing network traffic and system records. As mentioned by Saleem *et al.* (2023), the use of machine learning can analyse and categorise incoming URLs, which safeguards users against malicious sites and phishing scams.

2.3.2 Data Sources and Feature Engineering

When it comes to cybersecurity, machine learning's efficacy is heavily reliant on the quality and range of information sources it uses, as well as the precision of feature engineering. The data used by cybersecurity professionals to train machine learning algorithms comes from a wide variety of sources. Logs from the server, the firewall, and the Domain Name System (DNS) are all examples. As per the view of Sam-Shin *et al.* (2022), an all-encompassing picture of network actions and vulnerabilities is provided by a variety of data sources. Machine learning algorithms examine this information to spot trends and outliers that point to harmful actions; Particularly useful for training algorithms to recognise known attack patterns is historical attack data.

2.3.3 Future Trends in Machine Learning for Cybersecurity

Deep learning, a kind of machine learning, can automate threat detection in a more complex and reliable manner. To analyse complicated data, such as image-based threat detection and the comprehension of contextual information, researchers are increasingly turning to convolutional neural networks (CNNs) and recurrent neural networks (RNNs). Also, systems have been able to automatically adapt to new threats with the help of AI, allowing for faster reactions and less human involvement. Understanding user and system behaviour is at the heart of behavioural analysis, which is expected to grow in popularity since it allows for the early discovery of abnormalities. As cited by Umer *et al.* (2022), recognising that attacks often target human weaknesses, user-centric techniques put a premium on monitoring and safeguarding individual users. These methods are crucial as the complexity of attackers increases.

2.4 Detecting Malicious URLs using Deep Learning (DL)

According to Birthriya & Jain (2021), malicious URLs encompass hosting harmful content, including phishing, spam, and malicious ads. It is argued that a variety of approaches were implemented in the past, such as signature matching, blacklisting, and regular expression approaches, to detect malicious URLs. However, it is contended that these traditional approaches are largely ineffective in detecting malicious URLs. In recent times, the use of the Deep Learning (DL) technique, which is a sub-category of Machine Learning (ML), has gained considerable attention for detecting malicious URLs. Khonji et al. (2014) conducted a survey study on different types of phishing solution techniques. In this study, it was found that the application of deep learning offers the most promising solution. Mahdavifar & Ghorbani (2019) and Berman et al. (2019) add a conclusion by stating that Deep Learning (DL) has been successfully used to address a variety of cybersecurity concerns. These issues include detecting URL-based phishing, email spam, malware, and website defacement. The use of DL seems to be more effective than traditional strategies such as blacklists and heuristics in addressing these cybersecurity challenges.

In a study conducted by Catal et al. (2022), it is argued that there are different types of DL algorithms such as deep neural networks (DNN), convolutional neural networks (CNN), recurrent neural networks (RNN), and long short-term memory networks (LSTM). Nonetheless, it is contended that DNN and CNN are the most commonly used DL algorithms for the identification of fradulent URLs. Additionally, among different types of DL algorithms, DNN algorithms are claimed to be the most preferred. An important reason that makes DNN the most preferred algorithm is that it is based on the traditional Multi-Layer Perceptron (MLP) model, which is capable of learning complex associations between data. Another important reason is linked to the widespread application of MLP compared to other recently developed DL algorithms. CNN is identified as another of the most preferred DL algorithms used to detect malicious URLs. Notably, scholars in the past, such as Somesha et al. (2020); and Wei et al. (2020) have proposed a CNN-based model for the detection of malicious URLs.

In contrast, Ali et al. (2022) claimed that even though DL techniques offered effective tools for detecting malware and URL-based phishing, there are some important limitations linked to the use of DL and its algorithms. In this context, it is argued that the use of DL for the detection of malicious URLs requires a sheer volume of training data, such as malware samples, for DL algorithms. Hence, the biggest challenge is related to continuously updating the training dataset with recent malware. Another major challenge is related to the selection of the DL algorithm. In this context, it is claimed that there are different types of DL algorithms, but it is highly challenging to select the most effective algorithm.

2.5 Detecting Malicious URLs using Natural Language Processing (NLP)

According to Salloum et al. (2021), it has been argued that URL-based phishing has emerged as one of the rapidly growing cybercrimes in the last few years, which has a significant negative impact on individuals and businesses alike. It has become extremely important to find appropriate solutions to mitigate the cyber threat posed by URL-based phishing. It is stated that NLP and ML have gained considerable attention for detecting URL-based phishing. Peng et al. (2018) recommended a phishing detection model that involved the use of the NLP technique. The model was used to examine the email’s subject phrases to discover any chances of phishing based on the four key malicious guidelines, which included a malicious URL, malevolent questions, standard greetings, and a resounding tone. Notably, the email was classified as phishing if it contained a malicious URL or any other guidelines mentioned above.

The test demonstrated 95% precision and 91% recall. Similarly, Alhogail & Alsabih (2021) proposed a phishing detection model based on NLP and deep learning. In this study, it was found that the use of NLP enhances the categorisation of accuracy and performance compared to the DL technique, which has high precision and recall rates.

Another study conducted by Sahingoz et al. (2019) proposed a URL-based phishing detection model built on NLP-based features. In this regard, it has been argued that the use of NLP-based features allowed the detection of URL-based phishing, which was not discovered previously. Based on the test of the proposed model, it is further argued that the use of NLP-based features for detecting malicious URLs delivers better performance than the use of word vectors. Likewise, Samad et al. (2023) proposed an NLP based model for the detection of malicious URLs. In this study, it has been claimed that the use of an NLP-based model reduces the requirement for feature engineering and results in more efficient use of the datasets for detecting malicious URLs.

2.6 Behavioural Analysis in Malicious URL Detection

As stated by Abdul Haseeb-ur-Rehman *et al.* (2023), it analyses the data sent and the frequency with which the URL interacts with the systems it connects to. Security teams are notified when suspect patterns arise thanks to behavioural analysis, which monitors for deviations from "normal" behaviour. There are several benefits to using behavioural analysis. Because it does not depend on signs or traits, it is great at finding new kinds of threats. As per the view of Ahmed *et al.* (2023), this preventative strategy finds vulnerabilities that signature-based approaches overlook, such as zero-day exploits and novel attack methodologies. In addition, behavioural analysis lessens the possibility of false positives by taking into account the context and behaviour of URLs inside a given network. Accordingly, the term false positive in the context of cyber-security occurs when a security system wrongly detects and classifies legitimate websites or activities as malicious or forged (Kabir & Hartmann, 2018). However, there are several difficulties associated with using behavioural analysis. It calls for an all-encompassing knowledge of network and system behaviour, which are difficult to grasp in expansive and ever-changing settings.

Behavioural analysis is a proactive and adaptable defence against developing cyber threats that uses machine learning to evaluate the dynamic behaviour of URLs inside a network. The continuing development of behavioural analysis tools such as Exabeam and Darktrace and their integration into larger cybersecurity policies is crucial to staying ahead of hostile actors and successfully protecting digital assets and infrastructure as assaults grow more complex and elusive (Olaniyan et al., 2023).

2.6 Limitations and Challenges in Malicious URL Detection

Identifying harmful Uniform Resource Locators (URLs) is a crucial task in cybersecurity, and it comes with its own unique set of constraints and difficulties. As mentioned by Alaoui & El (2022), individuals responsible for safeguarding digital assets must continuously adapt to the evolving methods and strategies employed by cybercriminals. In this piece, the various obstacles and restrictions that hinder the complete implementation of malicious URL detection have been examined. Cybercriminals use a wide variety of techniques, such as encryption, phishing, and social engineering, among others, to avoid being caught by security measures (Bederna & Szadeczky, 2019). As illustrated by Alshingiti *et al.* (2023), malicious URLs are disguised as safe ones by using obfuscation, polymorphism, or cloaking techniques. This ongoing game of cat and mouse between attackers and defenders demands better, more flexible methods of detection. Malicious URLs are constantly evolving, with attackers changing parts of the URL to evade detection. Since traditional detection methods such as signature-based detection techniques are dependent on previously identified patterns and traits, keeping up with them is challenging. As illustrated by Aslan *et al.* (2023), high levels of complexity and ongoing model retraining are required for machine learning algorithms such as Natural Language Processing (NLP) and Deep Learning (DL) to detect these changes (Saleem et al., 2023; Afzal et al., 2021).

In the pursuit of pinpoint precision in the detection of malicious URLs, false positives are often introduced. Incorrectly identifying safe websites or user behaviour as dangerous is called a false positive. These put a burden on both security resources and the user experience by triggering false alarms and disrupting operations. As cited by Asmaa *et al.* (2023), maintaining a reasonable false-positive rate while maintaining high accuracy is a constant issue. Malicious URLs sometimes disguise themselves as genuine links in seemingly safe material. The meaning of such a URL depends heavily on the surrounding information. Identifying legitimate URLs from fraudulent ones now requires an examination of behaviour and context. Context-aware detection systems, however, are difficult and time-consuming to develop.

2.7. The Evolution of Cyber Threats

Cyber dangers continue to develop at a rapid pace, reflecting the dynamic nature of the modern technological world and the evolving goals of cybercriminals. As illustrated by Chen *et al.* (2023), this section delves into how cyber threats have evolved through time, illuminating the historical background and current trends that have moulded the character of cyber vulnerabilities. Cybersecurity is not a new idea, but it has developed considerably over the past several decades. Primitive viruses and worms, motivated mostly by curiosity or mischief, dominated the danger landscape in the early days of computers. When compared to the financial and political motivations behind most modern cyberattacks, these were on the lighter side.

As per Cai et al. (2018), there has been a considerable increase in cyberattacks in the past few years, which can be attributed to various factors such as the growing value of digital assets, increased access to IT skills, and advancements in IT technologies. Besides, politically motivated espionage and financial gain have contributed to the rapid rise of cyber-crimes such as internet fraud, digital resource stealing, and hacked accounts. Ejaz *et al.* (2023) also supported this view and stated that the rise in cybercrime such as hacking, hacktivism, and ransomware efforts is increasingly motivated by a desire to gain undue advantage in politics, money, or revenge.

Supply chain assaults, in which hackers target legitimate businesses to contaminate their goods, have increased in frequency in recent years. These assaults are very successful and difficult to detect since fraudsters enter several organisations indirectly using this method. Attacks of this magnitude and skill are becoming more common, as seen by high-profile cases like the SolarWinds hack. As per the view of Elsadig *et al.* (2022), the prevalence of state-sponsored cyberattacks is growing. Cyber-espionage, sabotage, and misinformation operations are conducted by nation-states, often employing sophisticated methods and technologies. Attribution of such assaults is difficult, adding another layer of difficulty to the task of discouraging state-sponsored terrorists. Ransomware attacks have become more common in the past decade, with scammers increasingly using encryption to hold firms' data for ransom.

2.8 Current State of Cybersecurity

Currently, a variety of defensive practices, such as improvements in software development, better personnel training, and improvements in maintenance practices, reduce the risk of cybercrime and increase the cost for cybercriminals. However, despite these defensive measures, cyber assaults have steadily increased, creating a complicated and ever-changing environment in the field of cyber-security today (Slayton, 2017). In an increasingly interconnected digital world, organisations, governments, and people are consistently faced with a major challenge: monitoring and tracking the variety of tactics and techniques implemented by cybercriminals. Cybersecurity attacks are increasing in both frequency and complexity. As per the view of Kumar *et al.* (2023), vulnerabilities in technology, human behaviour, and organisational structures are exploited by malicious actors ranging from lone hackers to well-funded criminal organisations and state-sponsored institutions. Ransomware, supply chain assaults, phishing, and very sophisticated espionage efforts are all examples of cyberattacks.

Ransomware assaults, which encrypt a user's data and hold it hostage until a ransom is paid, have recently spread like wildfire. Companies of all sizes are being attacked, from those in the healthcare industry to those in the essential infrastructure sector. Because of its financial benefits, ransomware-as-a-service has become more popular among hackers. As per Kim & Lee (2017), it is argued that security flaws in the IoT, such as weak authentication and authorization, and other essential infrastructure, are becoming more pressing issues. Inadequately protected Internet of Things devices can be used by cybercriminals as access points. Cybercriminals can take advantage of weak authentication and authorization to gain unauthorised access to IoT-based devices. For example, cybercriminals can launch attacks on critical infrastructure like power grids and healthcare systems, which can have severe implications for national security.

Cybersecurity measures have also improved with the advancements in AI technologies such as machine learning. As a result, proactive and multidimensional tactics that include cutting-edge technology, threat information exchange, and user education are increasingly being used by organisations and security professionals. As stated by Merlino *et al.* (2022), threats are now better identified and mitigated with the use of behavioural analysis and machine learning. It is crucial to increase people's and businesses' security awareness. However, there is a persistent skills gap in the cybersecurity industry, making it difficult to keep up with rising demand. This shortage of talent is a major obstacle for businesses.

2.9 Gaps in Current Research

Human-centric vulnerabilities are an area where knowledge is lacking in the field of cybersecurity. While much focus has been placed on the technological elements of cybersecurity, the importance of human behaviour is increasingly being acknowledged. There is a lack of research on the psychological aspects of cybersecurity, which include user actions, choices, and vulnerabilities to social engineering. Cybercriminals continue to focus on human-centric vulnerabilities, such as manipulating individuals and taking advantage of a lack of awareness of cybersecurity best practices through deceptive emails and malicious URLs. Hence, closing this knowledge gap is essential for establishing effective tactics and treatments to address these threats. There is also a significant knowledge gap about the identification and prevention of zero-day attacks. As cited by Rozi *et al.* (2022), a major difficulty is zero-day vulnerabilities, which are used by hackers before companies can provide fixes. Detecting and responding to these elusive threats requires academics to develop cutting-edge strategies that integrate threat intelligence, anomaly detection, and machine learning in light of the continually changing software and hardware ecosystem. Standardised measurements and approaches such as the NIST Cybersecurity Framework and Capability Maturity Model Integration are needed since quantifying cyber risk and endurance is a challenging and developing field (Rea-Guaman et al., 2017). Accordingly, standardised measurements and approaches provide frameworks and metrics essential for assessing and managing cyber risks.

There is also a lack of studies about user-centric security solutions that emphasise user satisfaction, availability, and agency. As per the view of Zhai *et al.* (2023), security experts and the general public find common ground if security solutions are designed to be both easy to use and effective. Finally, given the increasing significance of cloud environments in today's IT infrastructures, there is a pressing need for an in-depth study into the best ways to keep these spaces safe. Data sovereignty, shared responsibility models, and the continually shifting threat environment in cloud systems all need to be thoroughly investigated before effective security measures can be devised. Accordingly, data sovereignty refers to the control of information that circulates under national jurisdiction (Hummel et al., 2021). On the other hand, shared responsibility models define the duties of cloud service providers and users in managing new or modified threats to achieve a high level of security.

2.10 Summary

This chapter goes into the many methods, such as signature-based analysis, heuristic analysis, blacklisting, and machine learning, used to identify dangerous URLs. Signature-based analysis, which is extensively employed in antivirus software and intrusion detection systems, includes comparing URLs to a database of known dangerous patterns but has difficulty keeping up with developing threats. While blacklisting is effective at blocking known harmful websites, it has limitations when it comes to detecting new threats and keeping its lists current.

The domain, path, and query parameters of a URL are analysed using rule-based systems in heuristic analysis to spot anomalies or suspicious behaviour. This method is useful for spotting previously undiscovered harmful URLs without established patterns, but it can also lead to false positives. The use of machine learning to detect malicious URLs has become more useful, especially in recent years. Machine learning algorithms can adjust to the ever-changing cybersecurity environment by analysing large datasets to find patterns and traits indicative of harmful intent. Machine learning relies on high-quality datasets collected from diverse sources, including server logs, firewall data, and DNS records, as well as carefully engineered characteristics to improve the precision with which threats are identified. Improvements in deep learning, behavioural analysis, and the comprehension of user and system behaviour are on the horizon for machine learning's usage in cybersecurity.

Increased assaults using complex methods like social engineering and ransomware characterise the present status of cybersecurity. Hacks like the one at SolarWinds show that cybercriminals are increasingly focusing on the supply chain. There is continuing worry about state-sponsored cyber risks, such as those posed by espionage, sabotage, and misinformation efforts. Additionally, the use of AI and ML in attacks further complicates the state of cybersecurity. Exploring the vast body of literature on the detection and classification of rogue URLs provides valuable insights into the shifting and perpetually evolving aspect of digital dangers and the strategies used to counter them as the cybersecurity industry struggles with this persistent menace.

Chapter 3: Methodology

3.1 Research Philosophy

This chapter demonstrates the method for creating AI frameworks to recognise malicious URLs. Before conducting any investigation, it is important to identify the underlying assumptions of understanding reality. There are various philosophies for understanding the reality of research. In this research, the pragmatism philosophy is used. Pragmatists believe that research reality can only be understood if it supports action, i.e., if it involves approaching the problem identified in the research practically. Throughout the research, it is considered that research reality regarding the detection of malicious URLs can only be relevant if it supports action, i.e., if the AI algorithm can effectively detect them practically (Tsung, 2016). It is considered that there is no single reality of the subject; therefore, the subject must be observed from different viewpoints. Alternative philosophies, like positivism, are inappropriate, as according to this philosophy, only through factual knowledge reality can be understood. However, this subject is more than just obtaining facts and statistical analysis. Apart from facts, the analysis must support the problem of the detection of malicious URLs. Then again, another alternative, like interpretivism, is also inappropriate for this subject, as according to this philosophy, reality is socially created. Nevertheless, in this research, reality cannot be understood by only the social construct of the subject or by the interpretation of people's viewpoints.

3.2 Research Approach

In this research, an inductive approach is used. Inductive research is based on specific observations to reach a general conclusion. This research also flows from specific observations regarding malicious URLs to a general conclusion about the way of detecting them generally by using AI methods. This research begins with observing the ML models used to predict URLs. Based on this, the pattern of using ML models by using AI is reconsidered, and accordingly, a new AI model for the prediction of malicious URLs is made (Wilson, 2010). The research defines the research framework through an explanation of the approaches and methods used. The initial step of the research involves collecting and preparing the data. It involves obtaining URLs and resultant labels and specifying their malicious status. Furthermore, the null values were eradicated to enhance the integrity of the data. The attributes that contribute to recognising the malicious URLs are pulled out. The outcome was a dataset where the row indicates an exclusive URL defined by its attributes and related output tags specifying its category. The next step concentrates on creating a model by using neural networks to properly identify malicious URLs (Gómez & Muñoz, 2023). Different instance selection methods are used for comprehending the impact on the AI model's performance. In the next step, the performance of the AI model is evaluated for its effectiveness in detecting malicious URLs. Regarding the alternative approach, i.e., the deductive approach, it seems inappropriate for this research subject as it is based on evaluating any existing concept through hypothesis testing. Such a method is not suitable for this research, as there are no hypotheses to be tested in the study.

3.3 Research Design

In this research, exploratory design is used. The research is based on obtaining ideas and perceptions regarding the utilisation of AI for detecting malicious URLs. It helps to understand the current AI scenarios and how they are being used in such situations as phishing and other harmful online activity detection. It is believed that the use of AI can solve the problem of malicious online activities. In the exploratory design, the literature search method is used. It is the quickest and most inexpensive method to undertake the study. There is an incredible amount of information accessible in online libraries and commercial databases, which are used in research to gain insights on the subject. On the other hand, alternative designs like descriptive and causal are not suitable for this research. The descriptive design is useful for evaluating any population based on certain variables. In this research, there is no study of the population; thus, a descriptive design cannot be chosen. While the causal design is suitable for evaluating cause-and -effect connections, in this research, no such connection has been established.

3.4 Research Strategy

This study is conducted as archival research, which involves searching and extracting information from original archives, for instance, historical data on URLs. Credible sources like the Kaggle dataset are used to obtain accurate data and conduct research. It is a popular archive that provides reliable datasets used for study purposes. It provides evidence of URLs and increases understanding of malicious websites. Furthermore, in this research, the mixed method is used, namely, both qualitative and quantitative data are used. Regarding the quantitative approach, statistical analysis methods are used to evaluate the subject. Furthermore, concerning the qualitative approach, the data has been analysed descriptively through explanation. This helps to leverage the strengths of both methods and makes the research much more comprehensive and analytical. Furthermore, this also helps to observe the subject from both a subjective and objective viewpoint. The quantitative approach is based on using statistical methods for modelling the data and ML algorithms for assessing malicious URLs in a practical online environment (Collis & Hussey, 2014). On the other hand, the qualitative approach is based on a descriptive evaluation of malicious URLs to understand their characteristics and identify the risks. There are other alternative approaches, like the monomethod, which is not used because it will limit the research findings and will not make the research extensive. On the other hand, multi-methods are much more complex to be used in this research.

3.5 Data Collection and Analysis

In this research, only secondary data is used. As a part of data collection, a list of URLs is developed with various categories like phishing and malware from online sources. The Kaggle website is used for data collection. There are a huge number of URLs that are extracted from the website URL, indicating a unique address for recognising an online resource, like a webpage. Initially, it is planned to indicate the protocol, i.e., utilised to reclaim the item with HTTP. The IP address specifies the requested website, and the port stipulates the gateway that is required to be utilized to access the website component. The URL is the way to get the online object. There is a set of parameters that are utilised to identify values that permit other online activities.

Raw URLs are inadequate for the recognition of malignant URLs since they are unable to provide insightful information regarding the features of URLs that can support their classification. Therefore, feature mining is important because it alters raw URLs into quantifiable indicators that AI can efficiently process. Consequently, different input features are used for developing AI models, such as URL length, age of domains, and structural elements among others. The category of every URL, like phishing or malware, was utilised for the output feature (Abad, Gholamy, & Aslani, 2023). In this research, a stemming method is used, which is a normalisation method with a list of words in the URL converted into shortened root words to eradicate redundancy. This method is used to minimise the inflected words.

Tokenization is used in research to replace data with tokens, which retains all important information about the data. Ngram tokenization is also used in the research by dividing the texts into words and using N-grams for every word of a certain length. These methods are used for text analysis in the URLs. The count function is used in the research for evaluating the number of times specific values have occurred in the list of URLs, like length, letters, digits, special characters, abnormal URLs, and secure HTTP. It is important for the research since it involves an extensive amount of data, and the Python platform permits us to understand the frequency of the components in URLs.

3.5.1 Development of AI Models

The preparation of the dataset provided a rich basis for analysis. For dealing with computational ineffectiveness, instance selection is used. In the research dataset, which incorporates various groups such as benign, phishing, and malware, the analysis method concentrates on determining the borders that distinguish these categories. The feature engineering framework is used in the research, which assists in transforming raw data into features, which are later used for developing AI algorithms (Sun, 2021). It comprises the creation, transformation, extraction, and selection of features, which act as variables for the URL data.

3.5.2 Correlation Matrix

In this research, a correlation matrix is used to evaluate the relationships between the variables. This helps to evaluate the relationship between different attributes of URLs. Various URL aspects like URL type, letter count, digit count, special character count, abnormal URL, and secure HTTP are evaluated using the correlation matrix.

3.5.3 Confusion Matrix

The confusion matrix is used for analysing the data and evaluating the performance of the classification framework. It helps to reveal the positive and negative values generated by the AI model to identify and predict the data classification of defacement and phishing, among others.

3.5.4 Gradient Boost

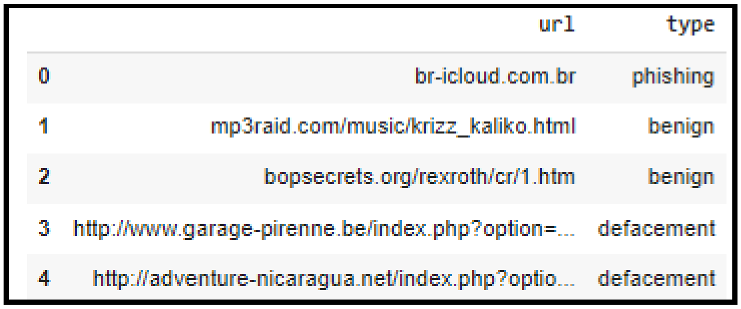
The gradient boost classification is used in the research to classify different URL classes based on precision, recall, and f1 score. Precision helps to understand the accuracy of the model’s positive predictions. Precision helps to examine the breakdown of actual positive cases that the model categorises as positive. For certain classes of URLs, such as malware, correctness is similar to the proportion of the URLs precisely recognised as malware against the entire URL projected as malicious. Recall evaluates the efficiency of the classification model in recognising every important instance from the dataset. It calculates the ratio of actual positive instances that the model properly recognizes. For the malware group, recall signifies the proportion of URLs that are properly categorised as malware compared to general real malware URLs. The F1 score is utilised for assessing the overall performance of the model. This score combines accuracy and recall to provide a balanced estimation of the two metrics in the ability of the AI model (Natekin & Knoll, 2013). In this way, this method helped to understand the capability of the AI model in discerning danger URLs.

3.6 Ethical Considerations

- In this research, it is ensured that any data taken from secondary sources is appropriate and genuine. No unauthorised data has been taken and used in the research.
- For maintaining data security, a password-protected machine is used for data storage.
- Proper academic representation is used in the research.
- Responsible information disclosure practices are followed in the research.
- Attempts have been made to reduce the bias and increase the impartiality of data collection.
- Ensured that participants received concise details regarding the study's purpose, document, participant information sheet, and permission form.
- To reduce the chance of being identified, complete anonymity was guaranteed.
- The emphasis was on voluntary involvement, with no pressure or improper influence.
- In adherence to the defined standards, the relevant ethics form has been signed and authorised.

Chapter 4: Result

Data Collection



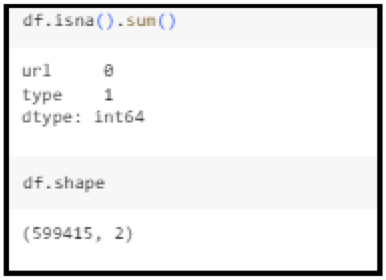
	url	type
0	br-icloud.com.br	phishing
1	mp3raid.com/music/krizz_kaliko.html	benign
2	bopsecrets.org/rexroth/cr/1.htm	benign
3	http://www.garage-pirenne.be/index.php?option=...	defacement
4	http://adventure-nicaragua.net/index.php?optio...	defacement

Figure 4.1: Showing the dataset

Using a dataset with a variety of URLs, the emphasis is on utilising advanced AI methods to identify potentially hazardous online links and then classify the different attack types associated with them. The collection of URLs covers a wide range of examples, from seemingly innocent ones like a link about music to more sinister ones like those connected to phishing and vandalism. By utilising characteristics such as the length of the URL, the existence of special characters, the age of the domain, and structural components, machine learning models can be trained to recognise trends suggestive of malevolent intent.

In addition to using widely used algorithms, the all-encompassing strategy also investigates the potential of deep learning techniques, including recurrent neural networks. Natural language processing (NLP) is incorporated to further improve the system's capability to examine the semantic material that is embedded in URLs. Real-time analysis ensures prompt recognition and response to possible hazards, while continuous learning mechanisms allow for adaptation to new threats. Proactive protection against the ever-changing panorama of cyber threats can be achieved through this framework's continual evaluation and refinement. The models' interpretability makes it easier to comprehend the reasoning behind a URL's classification as possibly harmful, which boosts trust in the effectiveness of the security system as a whole (Janet and Kumar, 2021). The dataset's varied array of cyberattacks is protected against using a seamless and comprehensive defensive mechanism due to integration with the current security architecture.

Data preprocessing

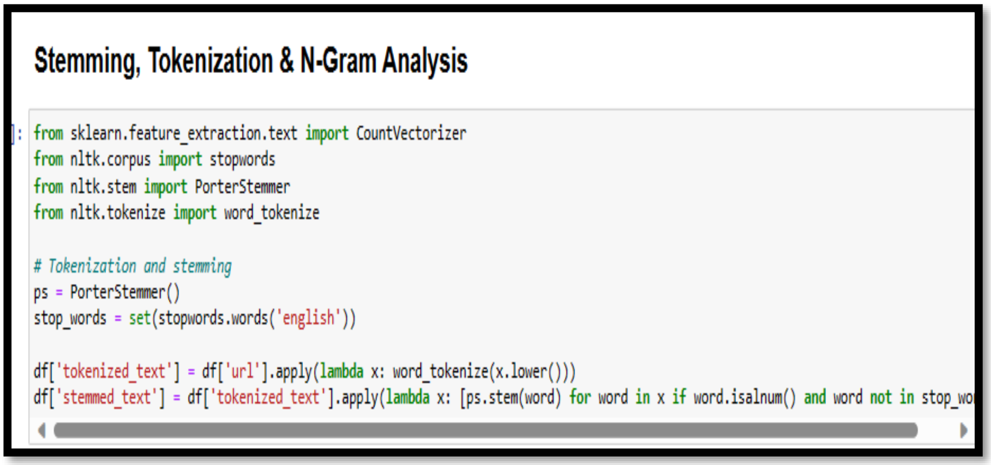


```
df.isna().sum()

url      0
type     1
dtype: int64

df.shape

(599415, 2)
```

Figure 4.2: Shape of the dataset of null values checking

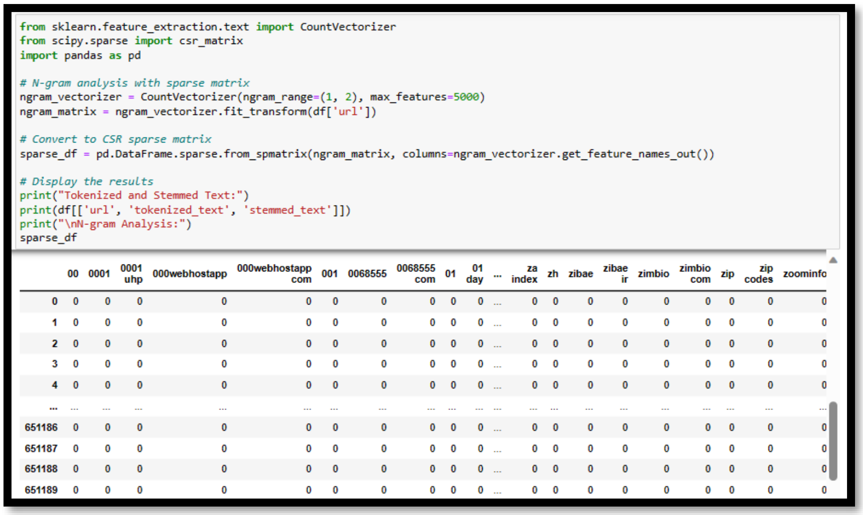
(Source: Acquired from Python Environment)

The shape of the dataset is represented here. This indicates that the dataset has 599415 columns and 2 rows. The null values are also checked as there are no null values present in this dataset.



Stemming, Tokenization & N-Gram Analysis

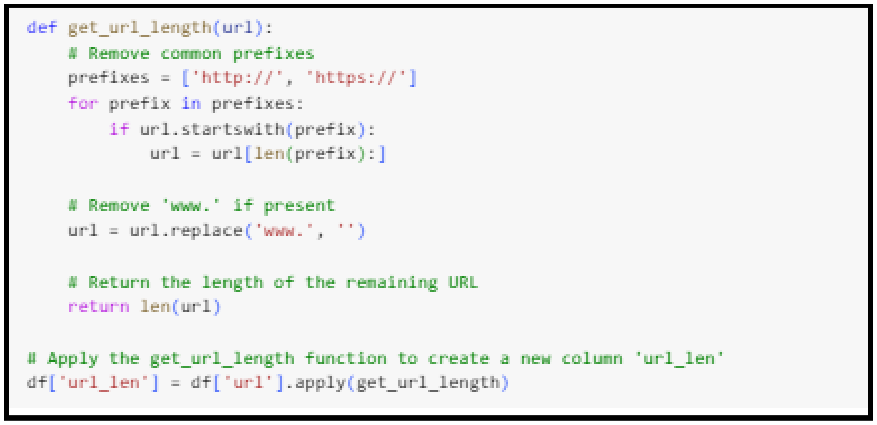
```
from sklearn.feature_extraction.text import CountVectorizer
from nltk.corpus import stopwords
from nltk.stem import PorterStemmer
from nltk.tokenize import word_tokenize

# Tokenization and stemming
ps = PorterStemmer()
stop_words = set(stopwords.words('english'))

df['tokenized_text'] = df['url'].apply(lambda x: word_tokenize(x.lower()))
df['stemmed_text'] = df['tokenized_text'].apply(lambda x: [ps.stem(word) for word in x if word.isalnum() and word not in stop_wo
```



```python
from sklearn.feature_extraction.text import CountVectorizer
from scipy.sparse import csr_matrix
import pandas as pd

# N-gram analysis with sparse matrix
ngram_vectorizer = CountVectorizer(ngram_range=(1, 2), max_features=5000)
ngram_matrix = ngram_vectorizer.fit_transform(df['url'])

# Convert to CSR sparse matrix
sparse_df = pd.DataFrame.sparse.from_spmatrix(ngram_matrix, columns=ngram_vectorizer.get_feature_names_out())

# Display the results
print("Tokenized and Stemmed Text:")
print(df[['url', 'tokenized_text', 'stemmed_text']])
print("\nN-gram Analysis:")
sparse_df
```

	00	0001	0001 uhp	000webhostapp	000webhostapp com	001	0068555	0068555 com	01	01 day	...	za index	zh	zibae	zibae ir	zimbio	zimbio com	zip	zip codes	zoominfo
0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
651186	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
651187	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
651188	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
651189	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0

Figure 4.3: Stemming, Tokenization & N-Gram analysis

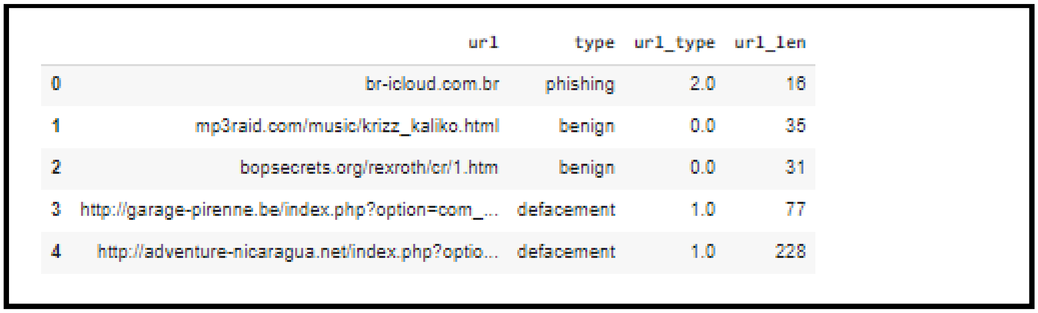


```python
def get_url_length(url):
    # Remove common prefixes
    prefixes = ['http://', 'https://']
    for prefix in prefixes:
        if url.startswith(prefix):
            url = url[len(prefix):]

    # Remove 'www.' if present
    url = url.replace('www.', '')

    # Return the length of the remaining URL
    return len(url)

# Apply the get_url_length function to create a new column 'url_len'
df['url_len'] = df['url'].apply(get_url_length)
```

Figure 4.4: Code for adding URL length

(Source: Acquired from Python Environment)

This adds a new column called "url_len" to the dataset and incorporates an additional function for determining URL length. The method efficiently determines the length of the final URLs by deleting common prefixes like 'http://' and 'https://,' as well as the 'www.' subdomain. This extra feature adds an essential component to AI models that are aimed at detecting dangerous URLs, resulting in a more sophisticated analysis that takes into account the structural characteristics of URLs within cybersecurity applications.

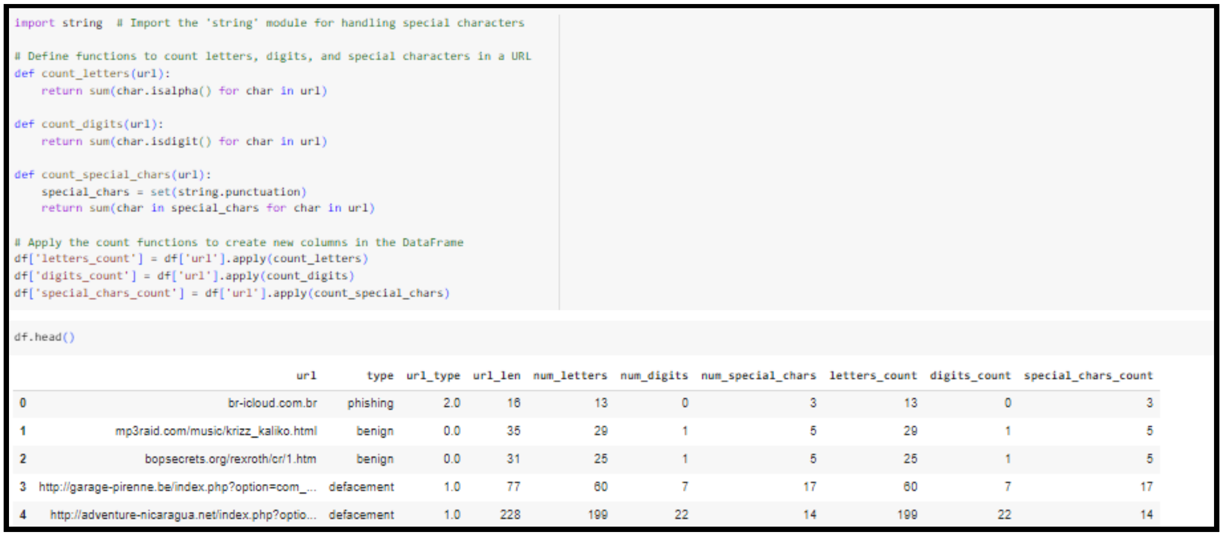


	url	type	url_type	url_len
0	br-icloud.com.br	phishing	2.0	16
1	mp3raid.com/music/krizz_kaliko.html	benign	0.0	35
2	bopsecrets.org/rexroth/cr/1.htm	benign	0.0	31
3	http://garage-pirenne.be/index.php?option=com_...	defacement	1.0	77
4	http://adventure-nicaragua.net/index.php?optio...	defacement	1.0	228

Figure 4.5: Columns after adding specified columns

(Source: Acquired from Python Environment)

The new URL type and the URL length columns are added to the dataset for effective analysis.



```
import string  # Import the 'string' module for handling special characters

# Define functions to count letters, digits, and special characters in a URL
def count_letters(url):
    return sum(char.isalpha() for char in url)

def count_digits(url):
    return sum(char.isdigit() for char in url)

def count_special_chars(url):
    special_chars = set(string.punctuation)
    return sum(char in special_chars for char in url)

# Apply the count functions to create new columns in the DataFrame
df['letters_count'] = df['url'].apply(count_letters)
df['digits_count'] = df['url'].apply(count_digits)
df['special_chars_count'] = df['url'].apply(count_special_chars)
```

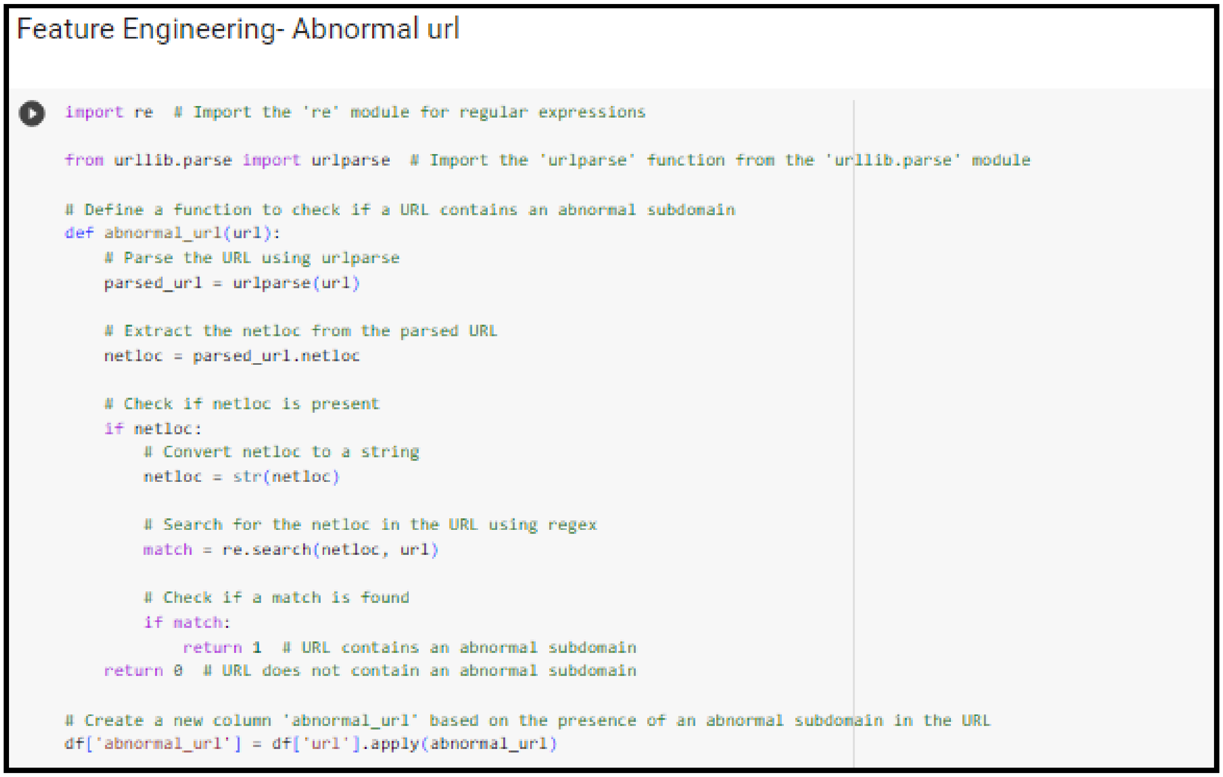
```
df.head()
```

	url	type	url_type	url_len	num_letters	num_digits	num_special_chars	letters_count	digits_count	special_chars_count
0	br-icloud.com.br	phishing	2.0	16	13	0	3	13	0	3
1	mp3raid.com/music/krizz_kaliko.html	benign	0.0	35	29	1	5	29	1	5
2	bopsecrets.org/rexroth/cr/1.htm	benign	0.0	31	25	1	5	25	1	5
3	http://garage-pirenne.be/index.php?option=com_...	defacement	1.0	77	60	7	17	60	7	17
4	http://adventure-nicaragua.net/index.php?optio...	defacement	1.0	228	199	22	14	199	22	14

Figure 4.6: Importing the count function

(Source: Acquired from Python Environment)

The code adds three additional columns to the dataset and enhances it through the incorporation of specialised functions in character analysis: 'letters_count,' 'digits_count,' and 'special_chars_count.' The aforementioned enhancements enable a more detailed analysis of URL structures concerning the identification of malicious activities. The methodical dissection of all the special characters, letters, and numbers in every URL provides insightful information on possible trends related to cyber threats. The basis for AI models that seek to identify malicious intent in cybersecurity apps through URL composition is strengthened by this detailed feature extraction (Sameen *et al.* 2020). The representation of the dataset after adding the specified columns is represented.



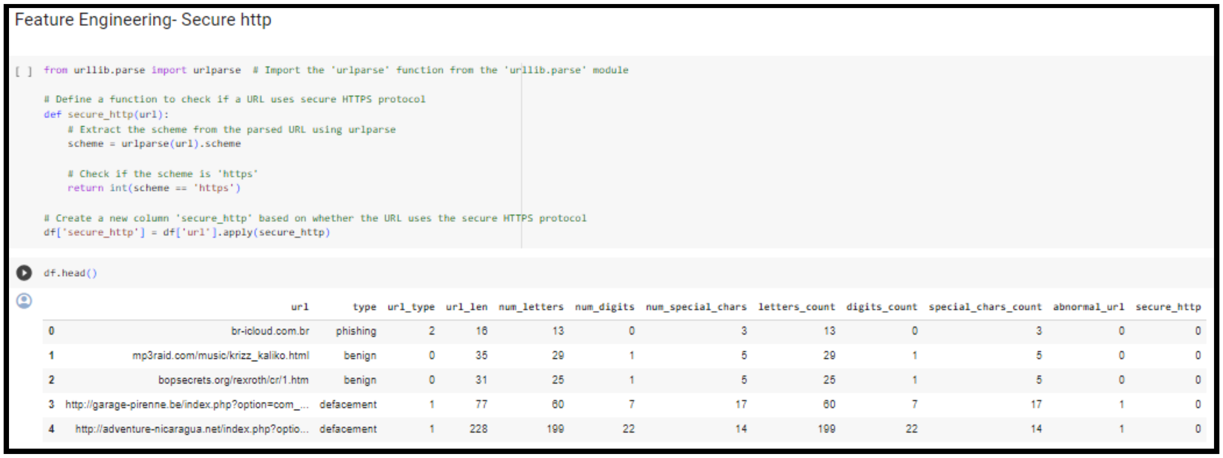
Feature Engineering- Abnormal url

```
import re  # Import the 're' module for regular expressions

from urllib.parse import urlparse  # Import the 'urlparse' function from the 'urllib.parse' module

# Define a function to check if a URL contains an abnormal subdomain
def abnormal_url(url):
    # Parse the URL using urlparse
    parsed_url = urlparse(url)

    # Extract the netloc from the parsed URL
    netloc = parsed_url.netloc

    # Check if netloc is present
    if netloc:
        # Convert netloc to a string
        netloc = str(netloc)

        # Search for the netloc in the URL using regex
        match = re.search(netloc, url)

        # Check if a match is found
        if match:
            return 1  # URL contains an abnormal subdomain
    return 0  # URL does not contain an abnormal subdomain

# Create a new column 'abnormal_url' based on the presence of an abnormal subdomain in the URL
df['abnormal_url'] = df['url'].apply(abnormal_url)
```

Figure 4.7: Feature Engineering of the Abnormal URL

(Source: Acquired from Python Environment)

By employing feature engineering to detect odd URLs, the code adds a crucial dimension to the dataset. The 'abnormal_url' function uses regular expressions and the 'urlparse' function to methodically analyse each URL and detect the existence of abnormal subdomains. The binary 'abnormal_url' column that is produced as a result improves the dataset's complexity and offers useful information for AI models that are designed to identify subtle trends that may indicate dangerous URLs. The continued development of safety features is made more robust as well as context-aware with the addition of this important feature.



Feature Engineering- Secure http

```
[ ] from urllib.parse import urlparse  # Import the 'urlparse' function from the 'urllib.parse' module

    # Define a function to check if a URL uses secure HTTPS protocol
    def secure_http(url):
        # Extract the scheme from the parsed URL using urlparse
        scheme = urlparse(url).scheme

        # Check if the scheme is 'https'
        return int(scheme == 'https')

    # Create a new column 'secure_http' based on whether the URL uses the secure HTTPS protocol
    df['secure_http'] = df['url'].apply(secure_http)
```

```
df.head()
```

	url	type	url_type	url_len	num_letters	num_digits	num_special_chars	letters_count	digits_count	special_chars_count	abnormal_url	secure_http
0	br-icloud.com.br	phishing	2	16	13	0	3	13	0	3	0	0
1	mp3raid.com/music/krizz_kaliko.html	benign	0	35	29	1	5	29	1	5	0	0
2	bopsecrets.org/rexroth/cr/1.htm	benign	0	31	25	1	5	25	1	5	0	0
3	http://garage-pirenne.be/index.php?option=com_...	defacement	1	77	60	7	17	60	7	17	1	0
4	http://adventure-nicaragua.net/index.php?optio...	defacement	1	228	199	22	14	199	22	14	1	0

Figure 4.8: Feature Engineering-Secure

(Source: Acquired from Python Environment)

The implementation of the 'secure_http' function provides a crucial security-oriented element to the dataset. It uses the 'urlparse' functions and the 'urllib.parse' modules to analyse every URL's scheme, paying particular attention to whether the secure HTTPS protocol is present. For AI models charged with detecting and classifying fraudulent URLs, the resulting binary indicator, 'secure_http' column, offers useful data. This improvement emphasises the value of secure communication channels throughout the continuous combat against online dangers, which is consistent with modern cybersecurity procedures.

EDA

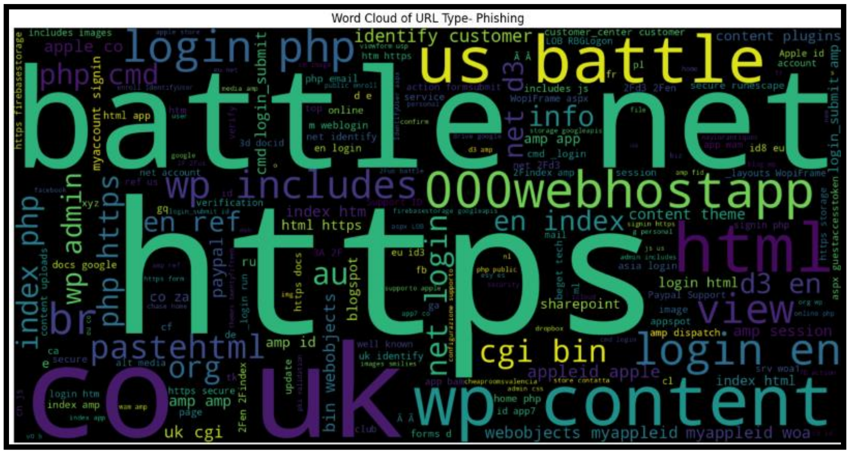


Figure 4.9: Word Cloud of URL phishing

(Source: Acquired from Python Environment)

This is the word cloud of the URL phishing. The most used words in these word clouds are https, battle, net, uk, and so on. A word cloud evaluation of phishing URLs might provide important phrases to help identify and categorise possible attacks. Important components include phrases that are used often, such as "https," which indicates the usage of secure connections, as well as words that are specifically mentioned, like "battle" and "net." The country code "uk" is included, which implies a specific location. This comprehensive linguistic analysis allows us to better understand the prevalent patterns linked to phishing assaults, which in turn assists in designing AI-based detection mechanisms that are more successful.

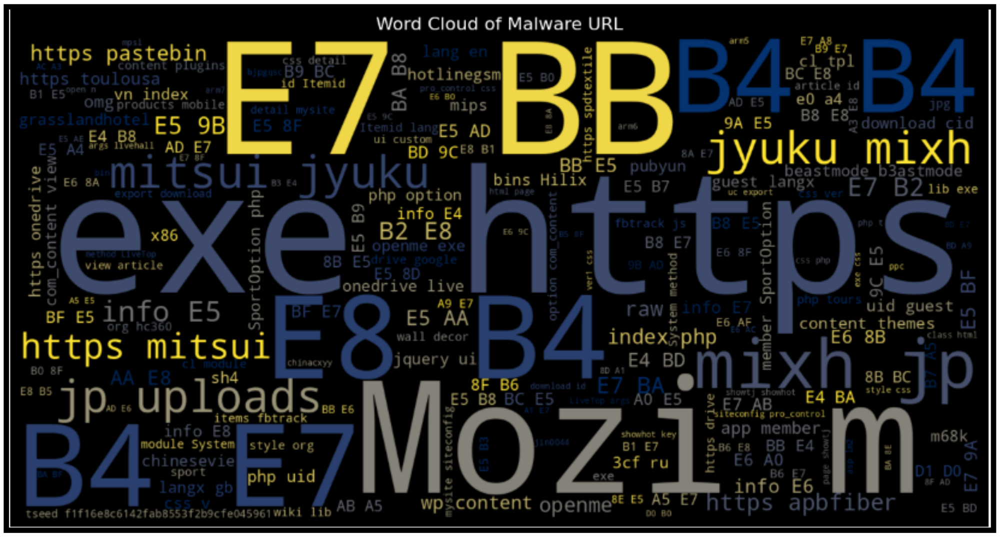


Figure 4.10: Word Cloud of Malware URL

(Source: Acquired from Python Environment)

The above is the representation of the word cloud for the malware URL. The most used words in these word clouds are E7, https, B4, E8 and so on.

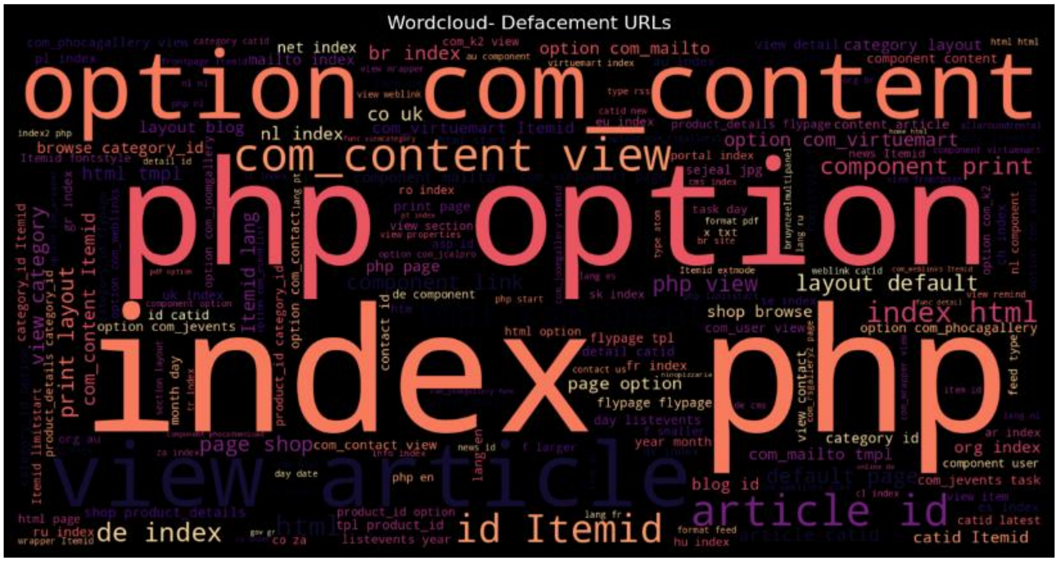


Figure 4.11: Word Cloud of the defacement URLs

(Source: Acquired from Python Environment)

It is the word cloud of the following defacement URLs. The most used words in this specific word cloud are index, php, option, com and so on.

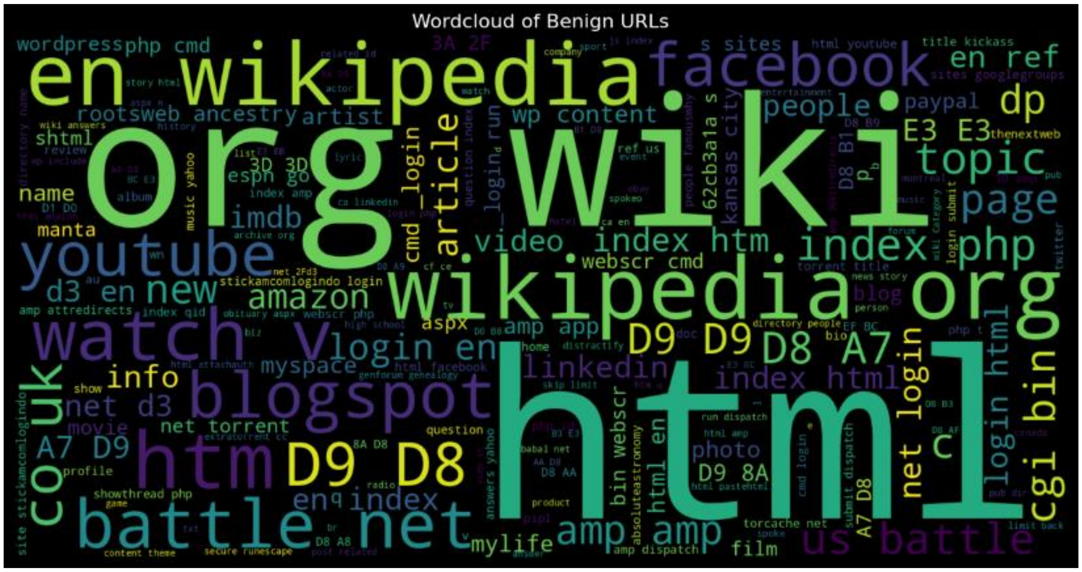


Figure 4.12: Wordcloud of Benign URLs

(Source: Acquired from Python Environment)

The picture above depicts the word cloud of the benign URLs. Here the most used words are html, wiki, org, facebook, and so on.

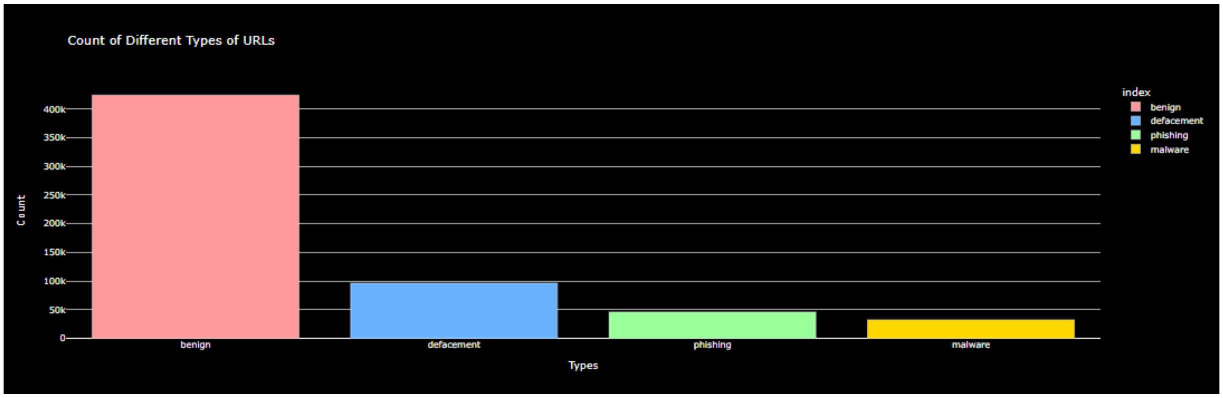


Figure 4.13: Count of the different types of URLs

(Source: Acquired from Python Environment)

The overall counts of the different types of the URLs are demonstrated here. The maximum count of the URL is benign which is about 400K and the less used URL is malware which is in count below 50K.

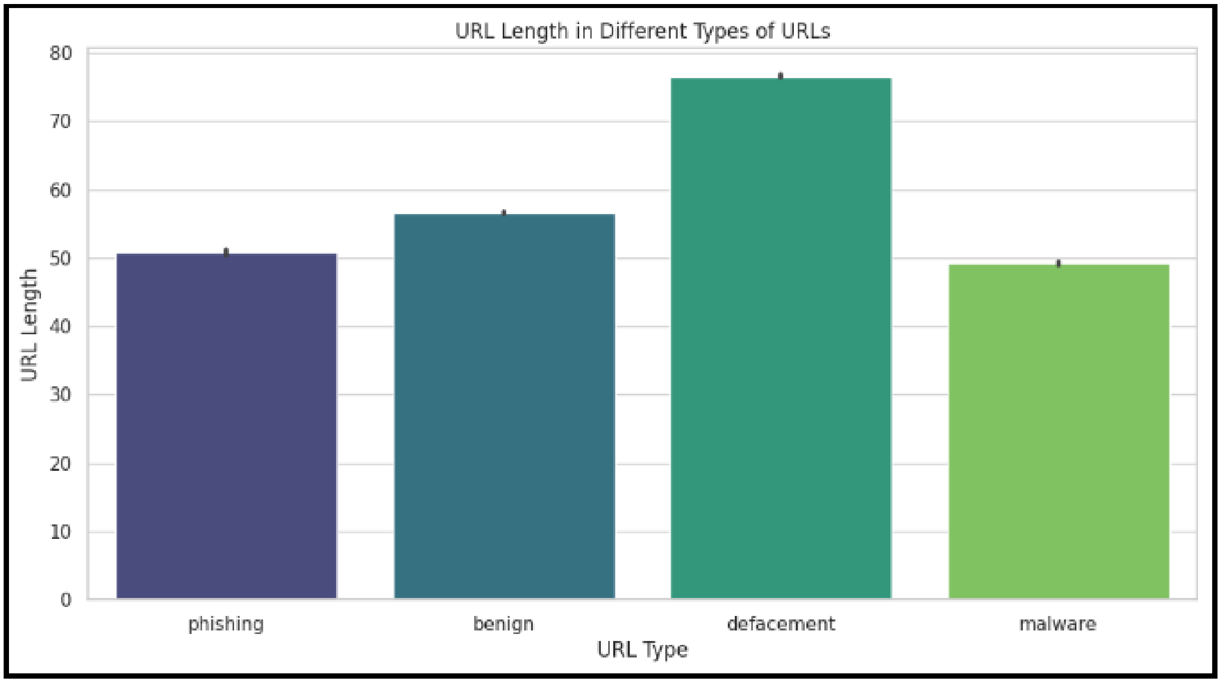


Figure 4.14: URL length in different types of URLs

(Source: Acquired from Python Environment)

The maximum length of the URL is identified in the defacement which is about 80. The minimum length is recognised in the malware type of URL which is below 50 counts.

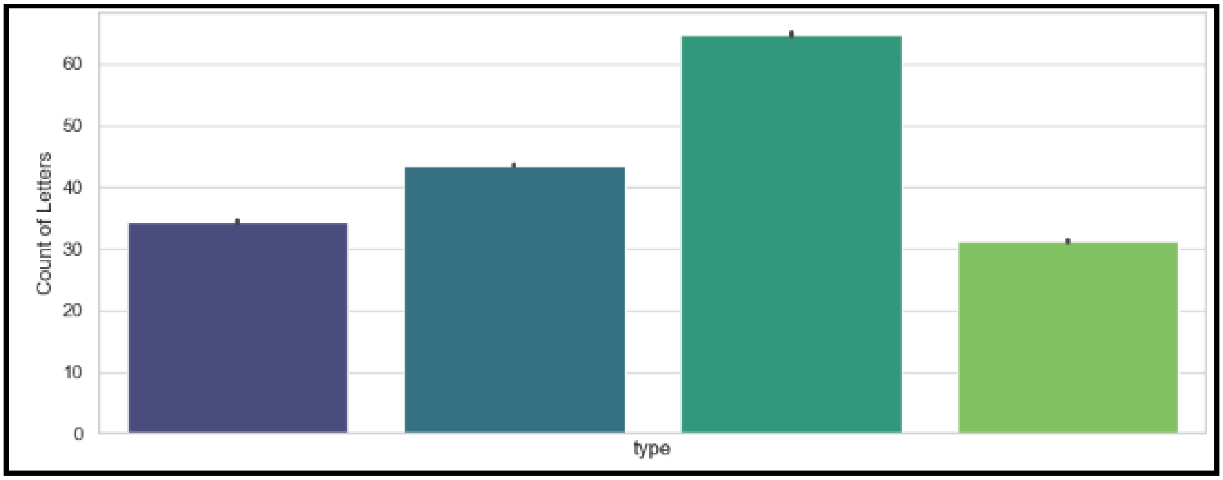


Figure 4.15: Count of Letters

(Source: Acquired from Python Environment)

The corresponding visual figure illustrates the overall count of the letters. The maximum counts are recognised in the type 3 URL and the minimum count is shown in type 1.

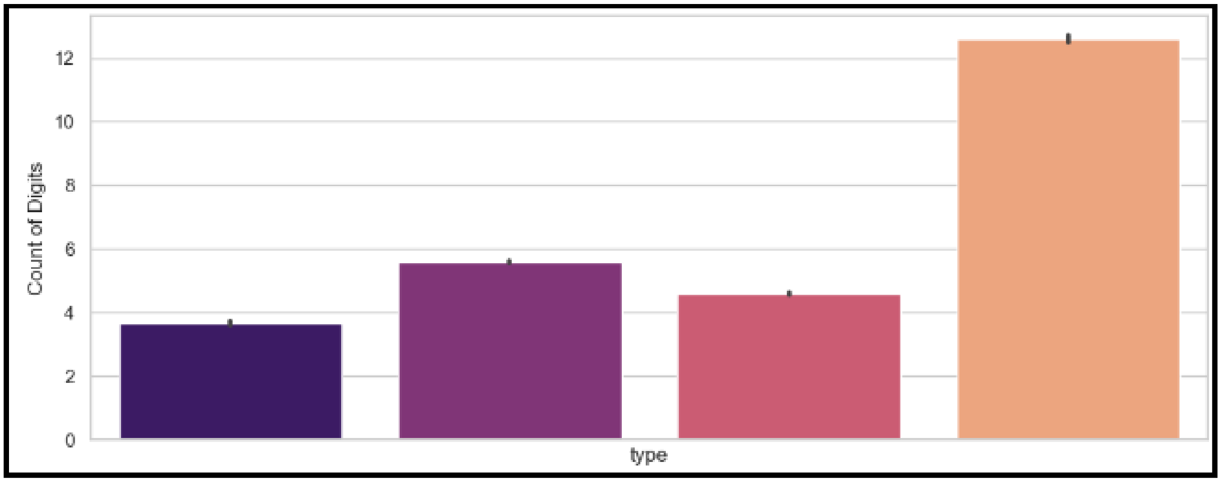


Figure 4.16: Count of the digits

(Source: Acquired from Python Environment)

In terms of the overall count of the digits, type 4 URL has the highest count, whereas type 1 shows the minimum count of the digits. The displayed figure, which shows the number of digits in URLs, offers an insightful look at the numerical makeup of various URL categories in the dataset. Particularly, URL type 4 has the most digits, indicating a unique number pattern linked to this particular category. On the other hand, URL type 1 has the fewest number of digits, suggesting a possible feature that sets it apart from other kinds. In the area of cyber threat identification, quantitative analysis serves as an essential component for machine learning models, advancing understanding of the structural variances across URL types.

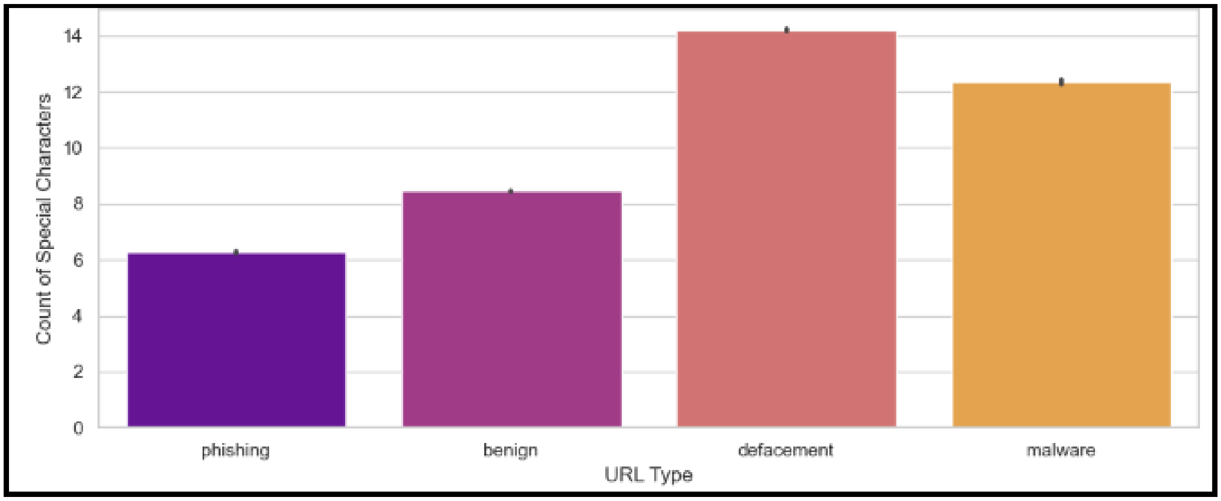


Figure 4.17: Count of the special characters

(Source: Acquired from Python Environment)

The count of the overall special characters is illustrated in this specific figure. The defacement has the highest count of the special characters, and the minimum number of characters is identified in the phishing.

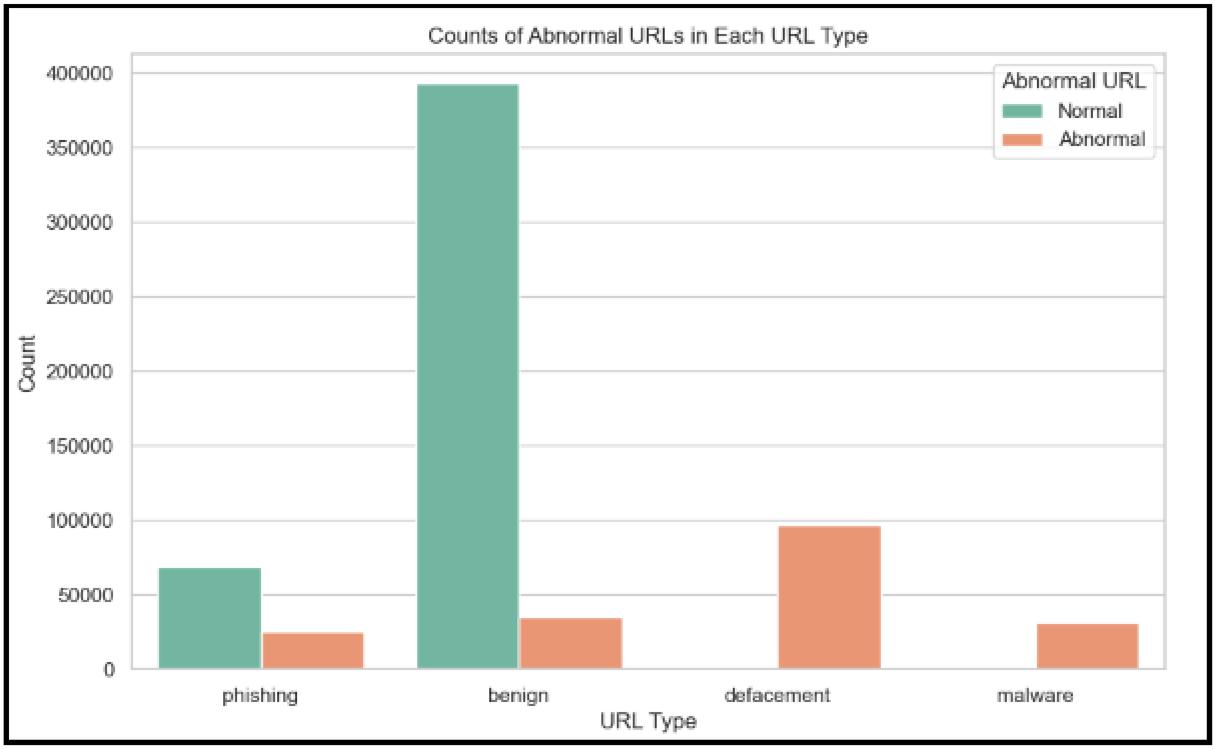


Figure 4.18: Counts of Abnormal URLs in each URL type

(Source: Acquired from Python Environment)

It is the demonstration of the overall count of the abnormal URLs in every URL type. The maximum count of normal URLs is realised in the benign type of URL. The count of this URL is near about 40000. The maximum count of the abnormal URL is seen in the defacement type of URL.

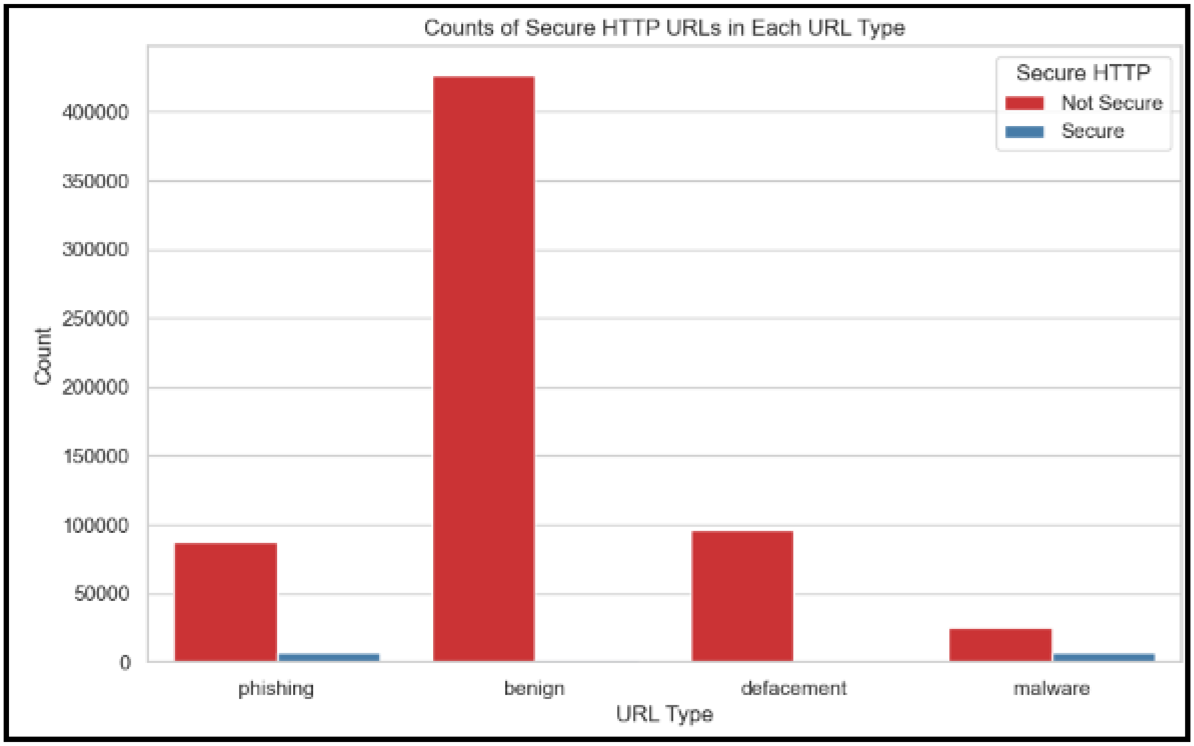


Figure 4.19: Counts of secure HTTP URLs in every URL type

(Source: Acquired from Python Environment)

The above illustration shows the overall counts of the secure and non-secure HTTP URLs. The maximum count of insecure URLs is in the benign type of URL. The most secure https are the phishing and malware type of URLs as there are no secure URLs in the benign and defacement types of URLs.

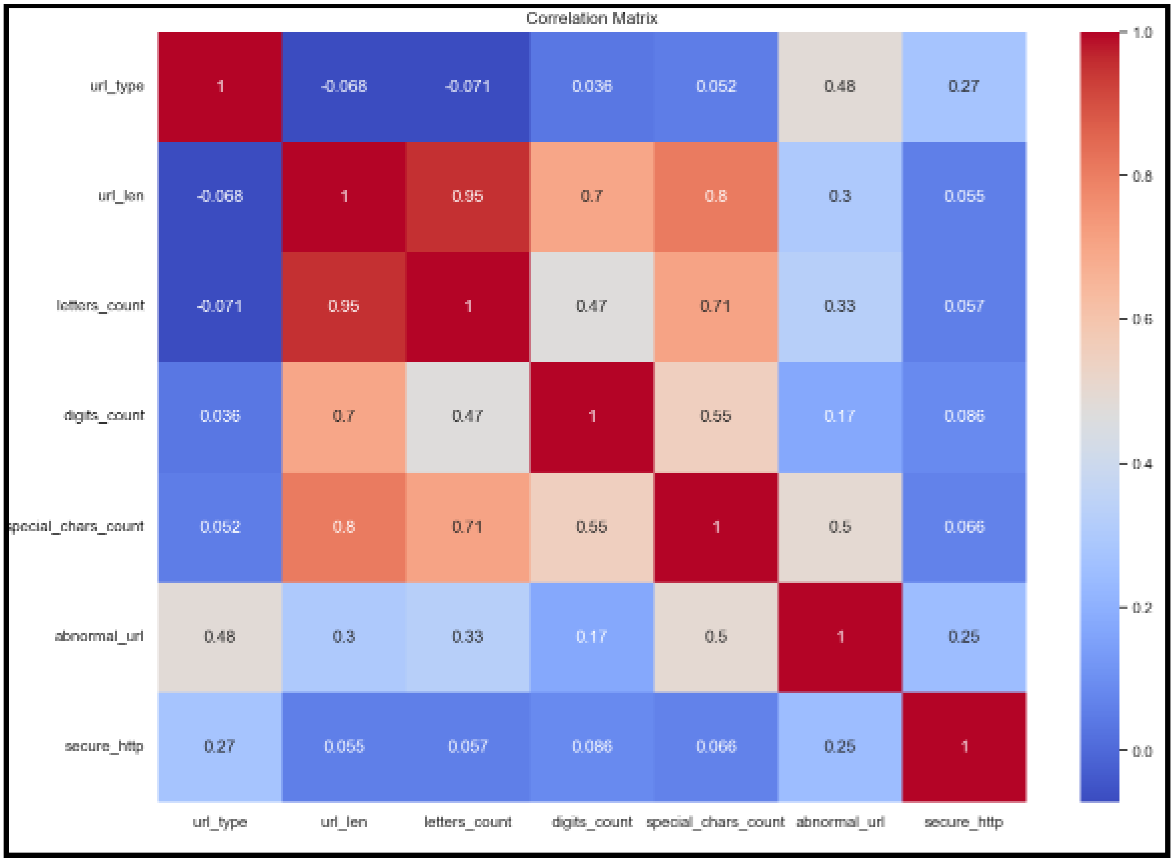
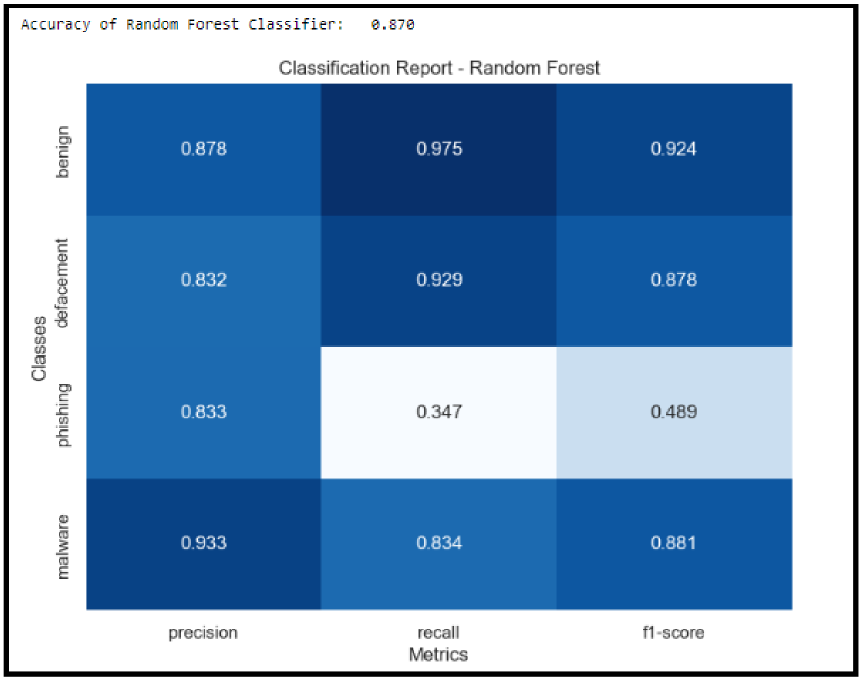


Figure 4.20: Showing the Correlation Matrix

(Source: Acquired from Python Environment)

The above image shows the correlation matrix of the specified attributes in the dataset. Some of the negative correlation values are -0.068, and -0.071. The positive correlation values consist of 0.55, 0.71, 0.036, and so on. As an essential tool for feature analysis, the correlation matrix that is being given provides a thorough understanding of the correlations between the various attributes in the dataset. A slight inverse association between similar qualities is shown by negative correlation values like -0.068 and -0.071. On the other hand, values of positive correlation, such as 0.55 and 0.71, indicate a more robust direct association between the related variables. The degree of the linear link is shown by these correlation coefficients, which assist in understanding how changes in one feature may affect changes in another. These kinds of insights are essential for model optimisation, feature selection, and dataset refinement for machine learning implementations. With its nuanced representation of attribute correlations, the correlation matrix improves the dataset's interpretability and facilitates well-informed decision-making during the development and evaluation of models for applications such as cybersecurity malware URL detection.

Model Evaluation



```
Accuracy of Random Forest Classifier:   0.870
```

Figure 4.21: Performance matrix of the Random Forest

(Source: Acquired from Python Environment)

The accuracy score of the Random Forest Classifier is about 87%. The Random Forest Classifier's performance matrix is displayed in the picture, and its impressive accuracy score of almost 87% highlights how well it can categorise URLs into various threat categories. This high accuracy is a result of the model's capacity to identify patterns throughout the dataset and generate accurate predictions. But to provide a thorough analysis, it's necessary to look more closely at the F1-score, precision, and recall metrics for each of the four different groups (malware, phishing, defacement, and benign). A thorough comprehension of the classifier's performance in accurately recognizing occurrences of each threat type, reducing false positives, and finding a balance between recall and accuracy can be obtained by analysing these metrics.

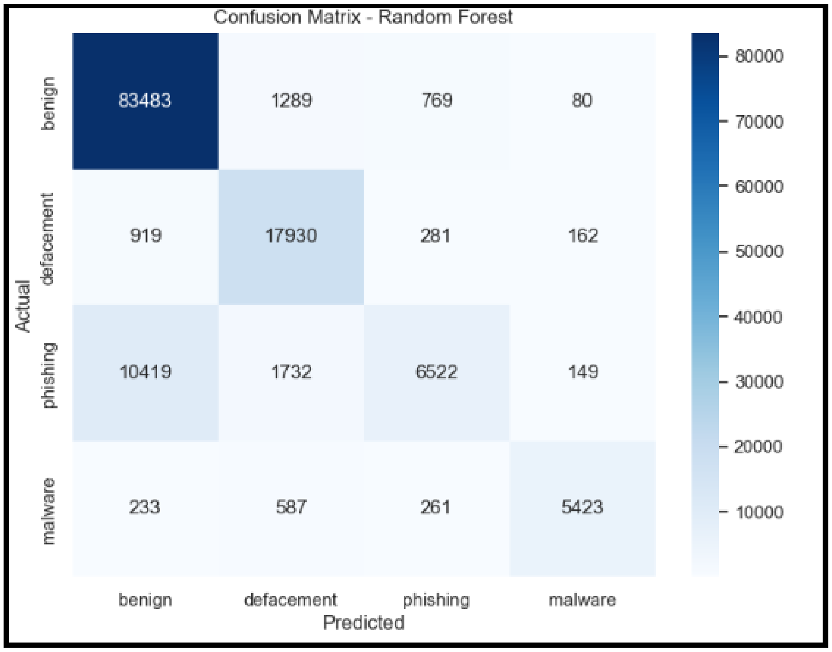
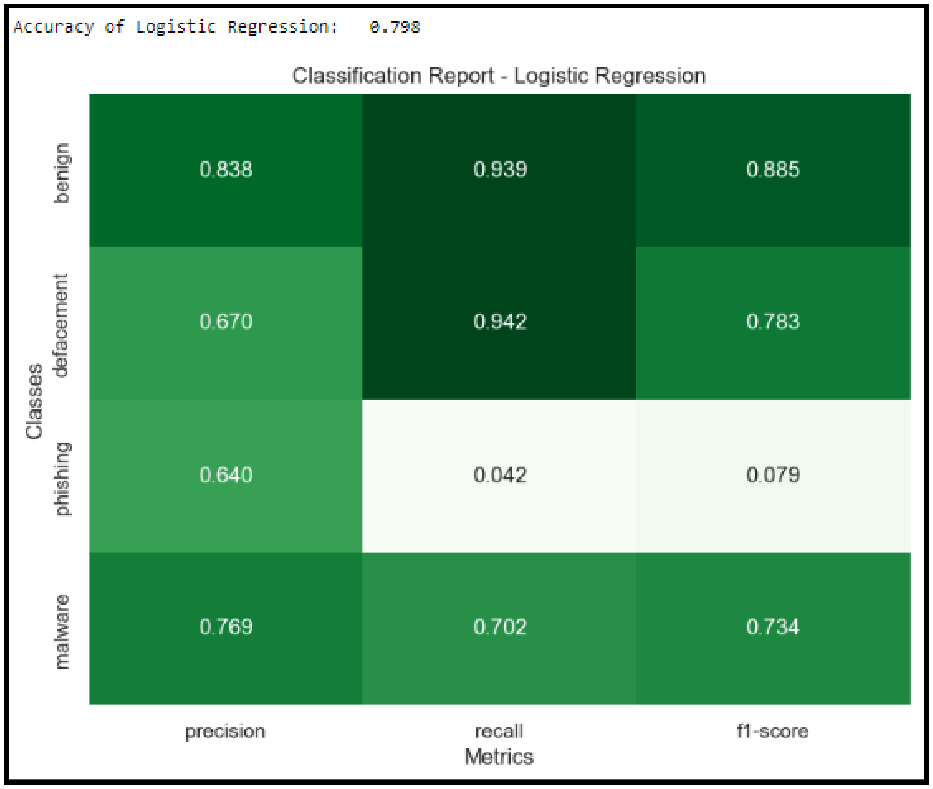


Figure 4.22: Confusion matrix of the Random forest

(Source: Acquired from Python Environment)

In the confusion matrix of the random forest, 83483 instances of benign were predicted correctly, while instances of other defacement, phishing, and malware were predicted wrongly. The mispredictions in the categories of defacement, phishing, and malware, however, indicate that the model has difficulty correctly identifying these types of incidents. The inaccuracy of these threat types' classifications highlights how difficult it is to discern minute patterns linked to various cyber threats. This indicates that the classifier needs to be further optimised and improved to increase its adaptability to the distinguishing characteristics of malware, phishing, and defacement URLs. By identifying certain regions that require work, the confusion matrix analysis facilitates the iterative procedure of model improvement (Johnson *et al.* 2020). It underlines the relevance of interpretability in guaranteeing the Random Forest Classifier's dependability in practical cybersecurity applications and highlights the necessity of continuous evaluation and modification to successfully minimise the constantly changing panorama of cyber dangers.



```
Accuracy of Logistic Regression:   0.798
```

	precision	recall	f1-score
benign	0.838	0.939	0.885
defacement	0.670	0.942	0.783
phishing	0.640	0.042	0.079
malware	0.769	0.702	0.734

Figure 4.23: Classification report of the Logistic Regression

(Source: Acquired from Python Environment)

The accuracy score of the logistic regression is about 79%. This accuracy score shows how well the model performs overall in terms of correctly predicting different threat types. However, a thorough assessment requires a more in-depth examination using precision, recall, and F1-score metrics for every class—malware, phishing, defacement, and benign. These metrics provide a detailed picture of the model's strengths and opportunities for improvement by revealing how well it minimises false positives and negatives. The accuracy score of the logistic regression is a useful criterion that directs further attempts to improve its predictive power in the ever-changing field of cybersecurity. Additionally, the classification report of the logistic regression, enhanced by metrics such as precision, recall, and F1-score, clarifies its complex performance in different threat categories. The accuracy score offers a general assessment, while the specific measurements shed light on particular advantages and potential improvement areas. For the model to become more predictive, it needs to be skilled at reducing false positives as well as negatives. As a guiding parameter, the accuracy score drives ongoing efforts to improve Logistic Regression's effectiveness in navigating the ever-changing cybersecurity landscape and strengthens its position as a useful tool in threat detection as well as mitigation techniques.

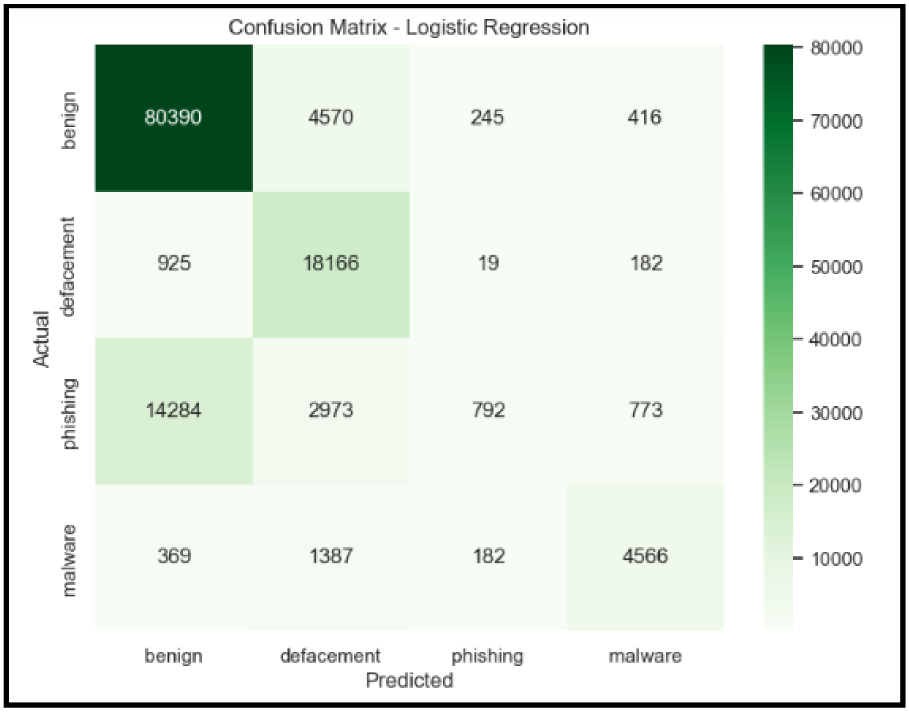
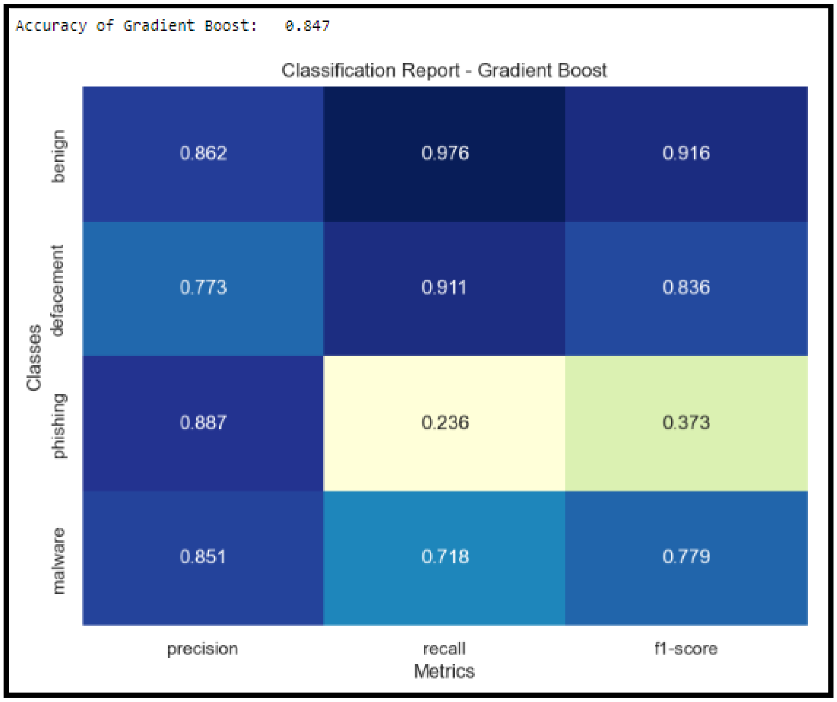


Figure 4.24: Confusion matrix of the Logistic Regression

(Source: Acquired from Python Environment)

In the confusion matrix of the logistic regression, 80390 benign was predicted correctly; another defacement, phishing, and malware were predicted wrongly. With 80,390 accurate predictions, the confusion matrix of the logistic regression model shows how well it performs at predicting benign instances. The mispredictions within the categories of defacement, phishing, and malware, however, show that the model is not entirely reliable in identifying these types of incidents. The statement highlights the complex difficulties in identifying minute trends connected to different cyber threats. The cases of misclassification emphasise the necessity of improving the sensitivity of the algorithm to distinguish between defacement, phishing, and malware URLs. To maintain the logistic regression model's effectiveness and dependability in practical cybersecurity applications, it is necessary to conduct ongoing assessments and modifications that take into account the ever-changing landscape of cyber threats.



```
Accuracy of Gradient Boost:   0.847
```

	precision	recall	f1-score
benign	0.862	0.976	0.916
defacement	0.773	0.911	0.836
phishing	0.887	0.236	0.373
malware	0.851	0.718	0.779

Figure 4.25: Classification Report of the Gradient Boost

(Source: Acquired from Python Environment)

Gradient Boost's accuracy score is approximately 84%, demonstrating the model's ability to predict outcomes across a variety of threat classifications. Nevertheless, a detailed analysis of the precision, recall, and F1-score metrics for every class—malware, phishing, defacement, and benign—provides a more nuanced view. These metrics show how well the model can reduce false positives as well as false negatives, which is important information for improving the model's predictive power. The excellent accuracy of the Gradient Boost provides a solid basis for its use in cybersecurity, demonstrating its adeptness in negotiating the complexities of various cyber threats. In addition, the metrics for precision, recall, and F1-score demonstrate how well the Gradient Boost predicts cases in all threat categories.

Significantly, the model's capacity to reduce false positives alongside false negatives emphasises its strong performance. The detailed analysis highlights how adaptable the Gradient Boost is at negotiating the complexities of cybersecurity, which makes it a useful tool for threat identification and mitigation. The detailed classification report highlights the model's dependability by providing a comprehensive view of its effectiveness in managing various cyber threats and directing continued work to maximise its prediction capacities for practical uses.

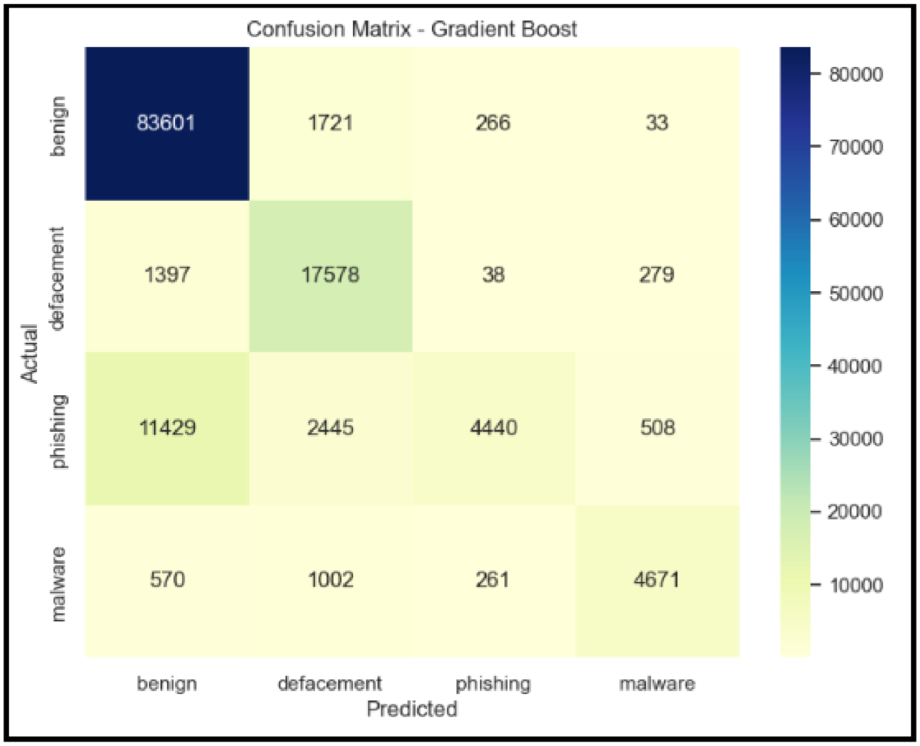


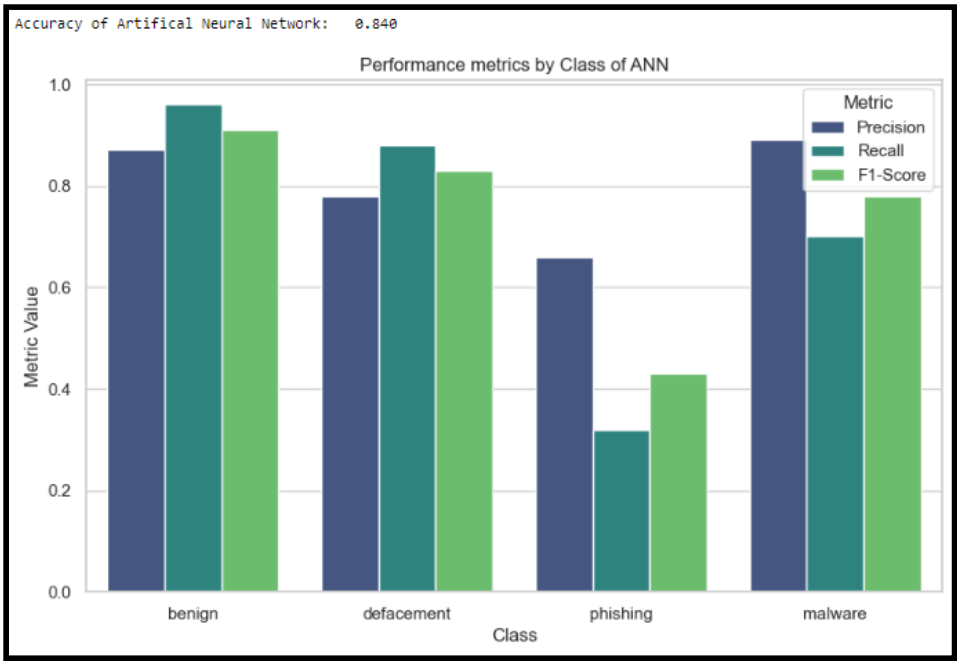
Figure 4.26: Confusion matrix of the Gradient Boost

(Source: Acquired from Python Environment)

This is identified in the confusion matrix of the gradient boost 83601 benign predicted correctly, other defacement, phishing, and malware predicted wrongly. With 83,601 right predictions, the Gradient Boost model's revealed confusion matrix highlights how good it is at anticipating benign events. But there are problems correctly categorising cases of malware, phishing, and defacement, as seen by the incorrect forecasts in these categories. This complexity demonstrates the model's effort to successfully identify minor patterns linked to different types of cyber threats (Aljabri *et al.*, 2022). The misclassifications highlight how much more model development is needed to increase sensitivity to the particular characteristics of malware, phishing, and defacement URLs. This dynamic evaluation highlights the model's ability to make a substantial contribution to threat detection as well as mitigation tactics, which corresponds with its ongoing adaptation to the changing cybersecurity scene.

Hyperparameter	**Value**
Number of Layers	3
Activation Function	ReLU
Output Activation	Softmax
Loss Function	Sparse Categorical Crossentropy
Optimizer	Adam
Epochs	10
Batch Size	32
Input Dimensions	X_train.shape[1] (Assuming X_train is the input feature matrix)

Hidden Layer 1	Units: 128, Activation: ReLU
Hidden Layer 2	Units: 64, Activation: ReLU
Output Layer	Units: 4 (for the 4 classes), Activation: Softmax



```
Accuracy of Artifical Neural Network:   0.840
```

Figure 4.27: Showing the accuracy of ANN

(Source: Acquired from Python Environment)

The accuracy of the ANN is about 84%. The benign class has the highest precision, recall, and f1 scores. The model's complex performance inside particular classes is revealed by a more thorough examination of precision, recall, and F1 scores. Among all categories, the benign class has the greatest precision, recall, and F1 scores, highlighting the ANN's exceptional capacity to recognise benign cases. This improved performance in the innocuous class demonstrates how well the model distinguishes between typical web material. Nonetheless, the overall accuracy highlights the ANN's proficiency in managing a variety of cyber threats. Continuous model adaptation and refining are necessary to maximise its ability to detect subtle patterns linked to malware, phishing, and defacement threats, as well as to ensure that it can adapt to the dynamic environment of cybersecurity.

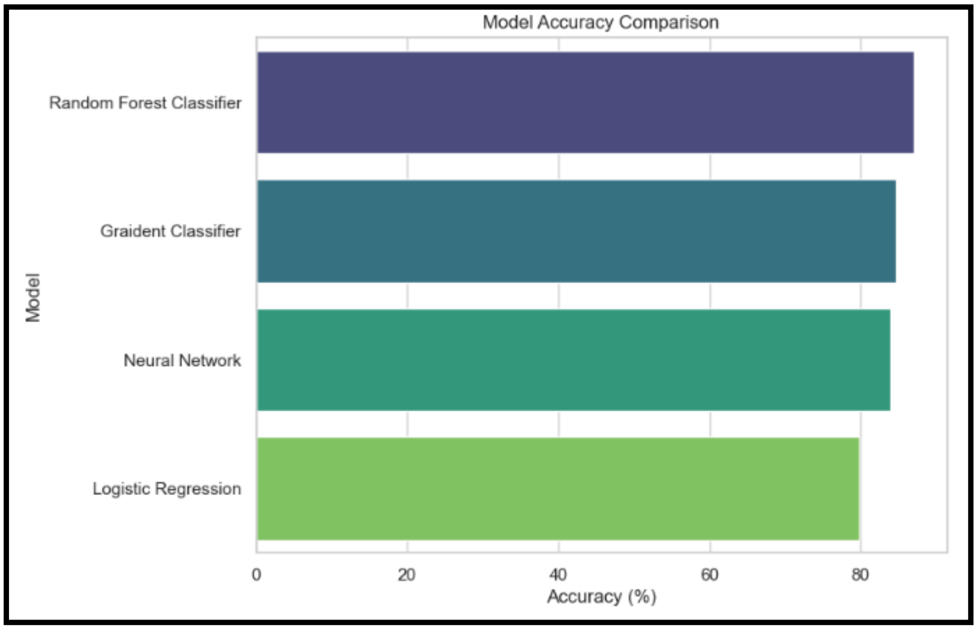


Figure 4.28: Comparison of the accuracy scores of implemented models

(Source: Acquired from Python Environment)

The contrasting graph of the respective constructed models is presented here. The accuracy score of the random forest classifier is the highest. The minimum accuracy score is identified in the logistic regression. A graph-based comparison examination of implemented models reveals clear performance trends. With the greatest accuracy score of every model, the Random Forest Classifier is the best performer. This demonstrates its outstanding ability to correctly categorise URLs into different threat categories. On the other hand, the Logistic Regression model has the lowest accuracy score, which indicates its predictive power is rather low. The range of accuracy scores demonstrates how differently the algorithms can identify patterns linked to malicious, phishing, defacement, and benign URLs.

Chapter 5: Discussion of Results

5.1 Introduction

The chapter presents the most important results from a study of secondary data that was gathered. The results of this investigation are contrasted with those from a literature review, which is also provided in this chapter. At the same time, this project aims to generate an effective system capable of detecting forged URLs and classifying attacks by integrating cybersecurity concepts with data analysis for enhanced threat identification accuracy. At the same time, the research topic that has been formulated divides this chapter into two parts. In addition, this chapter summarises the main issues raised during the discussion.

5.2 Methods for detecting URLs and classifying attacks

In the results section, the data gathering approach is laid out, with an emphasis on utilising powerful AI technologies to detect potentially dangerous connections on the internet and categorise the many forms of attacks linked to them. To instruct machine learning models to detect patterns that indicate malevolent intent, they use criteria including URL size, distinctive characters, domain age, and structural elements. The method also makes use of deep learning tools like NLP and recurrent neural networks to enhance the system's capacity to understand the semantic content included in URLs. To stay ahead of ever-changing cyber threats, applications have to be capable of assessing data in real-time and learn from their mistakes. The literature study delves into several methods for identifying illegal URLs. A technique that is often used, signature-based inspection, compares a URL to a database of known trends or signatures of malicious URLs (Mohammed et al., 2022). According to Nagy et al. (2023), while blacklisting is a simple approach that compares incoming URLs to lists of known dangerous websites, it has limitations in detecting novel, undiscovered risks. In contrast, the heuristic analysis uses rule-based algorithms to look for anomalies or suspicious behaviour in URL operations and characteristics. According to Ndichu et al. (2022), machine learning algorithms can analyse large datasets and find new threats, making them a useful tool for identifying malicious URLs. This study also emphasises the benefits of using models based on deep learning, such as CNNs and RNNs, to identify fake URLs (Uçar et al., 2019).

Examining the results side by side with the literature analysis reveals that both place stress on the power of machine learning techniques to identify dangerous URLs. Although the literature review discusses how machine learning algorithms have made use of URL structure, historical behavioural patterns, and content evaluation, the results section is more concerned with training models employing dataset features like domain age, specific characters, and URL size (Ndichu et al., 2022). To deal with ever-changing dangers related to cybersecurity, both parties acknowledge the significance of ongoing growth and change. On the other hand, the two components also disagree with one another. An improvement in the system's capacity to analyse the semantic data inherent in URLs is highlighted in the results section through the application of deep learning techniques, such as recurrent neural networks and natural language processing. According to the literature of Uçar et al. (2019), deep learning models such as CNNs and RNNs are useful for identifying fraudulent URLs, although natural language processing (NLP) is not mentioned anywhere about URL analysis. This disagreement points to a possible direction for further study: how to improve the identification of harmful URLs by integrating deep learning techniques, such as natural language processing.

The result highlights the need to use deep learning techniques and features extracted from the data set, which are two forms of sophisticated artificial intelligence, to detect relationships that might be harmful to users' online safety. Likewise, the study of the literature offers a summary of the methods currently used to identify malicious URLs. Although there are disagreements between the results and the literature review about the exact methods used (such as natural language processing and interpretability), they do agree that growing cyber threats necessitate constant learning and adjustment.

5.3 Malicious URL detection methods depending on attack types like malware and phishing

Findings from preprocessing data and EDA shed light on URL analysis and malicious activity identification. Supplementing the dataset with additional columns like "url_len," "letters_count," "digits_count," "special_chars_count," "abnormal_url," and "secure_http" allows AI models to better detect malicious URLs by including crucial structural properties. These supplementary elements enhance the foundation for artificial intelligence models used in applications related to cybersecurity and lead to a more thorough analysis (Sameen et al., 2020). Word clouds, digit counts, special character counts, anomalous URLs, and secure HTTP URLs are just a few of the dataset features examined in the EDA section. Insightful developments and patterns concerning cyber dangers, safe communication routes, and phishing assaults can be uncovered by these assessments. Word cloud evaluation, for instance, draws attention to key terms like "https," "battle," and "net" that may aid in the detection and classification of potential assaults. Various URL types may be better understood by looking at their digits and distinctive character counts, which reveal their numerical makeup and unique traits. Improving security measures that are both strong and sensitive to context can be achieved by the examination of both unusual and secure HTTP URLs (Sameen et al., 2020). The literature study delves into the topic of identifying harmful URLs using deep learning (DL) and natural language processing (NLP) methods. In comparison to more conventional methods that include blacklists and heuristics, deep learning (DL) approaches, especially CNN and DNN, offer promising remedies for cyber security issues like URL-based detection of phishing, spam email identification, identification of malware, and website vandalism identification (Birthriya & Jain, 2021; Mahdavifar & Ghorbani, 2019; Berman et al., 2019).

However, there are a few problems with DL approaches that are brought up in the literature review as well. These include the need for a huge amount of training data and the difficulty of choosing the best algorithm. Challenges remain in the application of DL to handle cyber threats, and training datasets must be continuously updated with the latest malware (Ali et al., 2022; Benavides et al., 2020). Additionally, natural language processing methods have improved the accuracy and performance of URL-based phishing detection. According to Salloum et al. (2021), Alhogail & Alsabih (2021), and Samad et al. (2023), using characteristics based on natural language processing decreases the requirement for feature engineering and enhances the utilisation of datasets for identification. The results show that the data pretreatment and EDA approaches are in line with the improvements in DL and NLP for detecting malicious URLs, as seen in the literature review. The significance of linguistic and structural trends in detecting cyber risks is reflected in the examination of URL features and the inclusion of additional parameters. When implementing AI models to detect harmful URLs, it is important to keep in mind the difficulties highlighted in the literature study. These include the necessity of constantly updating data and selecting algorithms. Thus, the findings and literature study work hand in hand to highlight the capability and promise of employing cutting-edge methods to strengthen cybersecurity protocols. In addition, based on the data from the result chapter, the analysis of the four models revealed that they were not all equally good at identifying various cyber risks; the top models were Random Forest Classifier (87% accuracy), Gradient Boost (84% accuracy), and ANN(84% accuracy), and Logistic Regression (79% accuracy). Nevertheless, errors in categorization were still detected, particularly when it came to identifying URLs associated with vandalism, phishing, and malware, according to the confusion matrix of all four models.

On the other hand, depending on the outcomes of the literature analysis, DL and NLP approaches have become increasingly popular for identifying dangerous URLs, providing better results than conventional methods. Still, there are a few obstacles to overcome, including picking the best algorithm and regularly adding new viruses to the training dataset. The study found that DNN and CNN are the most frequently used DL algorithms and that ML may be made more successful at identifying dangerous URLs with the help of NLP-based characteristics. Although the models showed encouraging accuracy levels in their evaluations, the literature research revealed certain gaps and difficulties in utilising DL and NLP approaches to identify harmful URLs. Consequently, to make the models and procedures more sensitive to the unique traits of various cyber threats, they must undergo constant modification and refinement.

5.4 Summary

To identify and categorise harmful URLs, the results section and literature review both emphasise the utilisation of artificial intelligence technologies, such as deep learning and language processing algorithms. To detect harmful URLs, several techniques are employed, such as blacklisting, heuristic analysis, signature-based evaluation, and machine learning algorithms. Although opinions vary on specific approaches, the study acknowledges that the ever-changing nature of cyber threats necessitates ongoing development and adaptation. Logistic Regression, Random Forest Classifier, Gradient Boost, and Artificial Neural Networks were the best models for detecting cyber threats. Results from DL and NLP techniques are encouraging; however, there are obstacles to overcome when choosing algorithms and updating datasets. However, for better cyber threat identification, continuous improvement is essential.

5.5 Evaluation of Potential Beneficiary

The feedback has helped me to evaluate the project's impact and effectiveness. It is also useful in focusing the project's emphasis and verifying that it solves real-world cybersecurity concerns. Furthermore, the investigation provided detailed insights into the reported success of the proposed strategies for detecting malicious URLs and categorising attack types. The potential recipient suggested a desire for the technical features of the AI approaches used, and remarks indicated appreciation for the project's potential contributions to the area.

In addition, input on the accessibility and feasibility of applying the proposed solutions gave important considerations for their practical use and adoption in real-world cybersecurity environments. The analysis provided helpful feedback and ideas for improvement, emphasising areas in which additional improvement and explanation could be advantageous.

The insights gained contribute to the project's continuing enhancement, ensuring that it not only meets academic standards but also addresses the practical challenges that cybersecurity professionals and organisations face in dealing with the constantly evolving environment of cyber threats. By considering the potential beneficiary's comments and concerns, I hope to improve the project's effectiveness, robustness, and broad contribution to the subject of cybersecurity.

Potential beneficiary feedback, participant information sheet, and consent form are carefully organised and properly stored in the University F drive.

5.6 Discussion on the Project Timelines

The dissertation schedules were systematically prepared to ensure a structured and continuous approach to completing critical objectives. The timeline included several steps, such as the preparation phase, which included a literature review, proposal development, data collection, model development and optimisation, model deployment, and the final write-up. Significant progress has been accomplished by the timeline, with each step adding to the overall progression of the research.

However, it is essential to acknowledge unforeseen challenges throughout its development. The preparatory step, particularly the literature review, had problems accessing relevant literature, resulting in a longer evaluation timeline. Unexpected problems with data collection required a schedule change to allow for a complete data evaluation. Technical problems during model building required an extensive debugging and refining stage. Delays in obtaining ethical permissions impacted multiple stages, forcing a revised timeline to incorporate appropriate ethical processes. Writing issues and unexpected backlogs in document completion affected the final write-up stage, creating a new deadline for comprehensive editing and finalisation. Despite these hurdles, adaptation and strategic adjustments were vital to achieving the dissertation's critical objectives.

In evaluating my project's deliverables, I carefully evaluated numerous critical factors. The literature evaluation is notable for its comprehensiveness, which provides a solid foundation for the study questions and objectives. The robustness of the research design and methodology is well aligned with the research objectives, with any changes made at this point directed at improving the reliability of the study. The fineness of data collection and accurate approach to analysis are clear, with initial results indicating acceptance of the research goals. Concerning writing outcomes, the project is significantly advanced, with drafting sections and chapters being constantly cleaned, resulting in the formation of a structured and logical narration. Reflecting on the schedules, while adaptations were required due to unexpected challenges, I am confident that these modifications were critical to the project's overall coherence and excellence. The timing changes not only solved problems but also allowed for a more in-depth examination of some aspects, which considerably increased the study's validity.

Chapter 6: Conclusion and Recommendations

6.1 Conclusion

The goal of this dissertation was to develop a system that could identify attacks and detect fake URLs by combining data analysis with cyber security principles for better threat detection. The study utilised the pragmatism philosophy to provide thorough and insightful results regarding an efficient system that can identify fake URLs. Furthermore, quantitative and qualitative methods were both utilised in the present study. The research used secondary data collected from the Kaggle platform. In this regard, a list of URLs classified as phishing or malware is compiled from multiple web sources as part of the data-collecting process.

It has been identified in the project that the literature study reveals several methods used to identify malicious URLs. One common method is signature-based inspection, which compares URLs to a database of known patterns or signatures of malicious URLs. Nevertheless, it has limitations when it comes to identifying new and unexpected dangers. The effectiveness of blacklisting, which involves matching URLs with a database of known dangerous websites, is questionable when it comes to detecting new threats and keeping the blacklist current. Although heuristic analysis can detect anomalies or suspicious activity by analysing URL traits and behaviour, it is not immune to false positives. The dynamic field of cybersecurity is one area where this flexibility is highly prized. By analysing the structure and content of a website, deep learning models like RNNs and CNNs have demonstrated remarkable accuracy in identifying bogus websites. Unfortunately, attackers' evasion tactics and the speed with which harmful URLs are altered continue to make it difficult to detect misleading URLs. Utilising sophisticated AI technologies to detect malicious web links and categorise various forms of assault is an integral part of the data collection process. To detect patterns suggestive of malevolent intent, machine learning models that encompass deep learning methodologies and natural language processing (NLP) are trained utilising characteristics including URL size, unique characters, domain age, and structural elements. More faith in the efficacy of the security system is fostered by the models' interpretability. Integrating these approaches into current security structures allows for the establishment of a thorough and smooth defensive system, protecting against a wide variety of assaults.

According to the study's findings, both the literature review and the study's results provide new insight into the topic of malicious URL identification utilising DL and NLP approaches. Conventional methods for identifying harmful URLs, such as signature matching, blacklisting, and regular expressions, are mostly useless. However, when it comes to cyber security issues like URL-based phishing, email spam identification, virus detection, and website defacement detection, DL and NLP approaches have demonstrated encouraging outcomes. Some DL algorithms, such as CNN and DNN, have recently come under scrutiny for their potential applications in identifying fraudulent URLs. Preference is given to DNN algorithms that learn complicated correlations between data based on the classic Multi-Layer Perceptron (MLP) model. Similarly, CNN algorithms have seen extensive usage in the fight against rogue URLs. The results produced by these DL methods outperform those of more conventional approaches, such as heuristics and blacklists. The application of DL and NLP to the problem of malicious URL detection is not without its difficulties, though. Some of these obstacles include choosing the best DL algorithm, constantly updating the training database with new malware samples, and a huge amount of training data. Regardless of these obstacles, DL and NLP methods are useful for improving machine learning models and finding and preventing harmful URLs.

The features of malicious URLs can be better understood thanks to the findings of data preparation and exploratory data analysis (EDA). When looking for malicious URLs, it is vital to consider features like URL length, character evaluation, unusual subdomains, and secure HTTP protocols. Improved artificial intelligence algorithms that can detect and categorise malicious URLs are a direct result of these findings. In addition, the study of four models, ANN, Random Forest Classifier, Logistic Regression, and Gradient Boost used to classify URLs as threats sheds light on their relative merits and suggests ways forward for cybersecurity research. The Random Forest Classifier was the most accurate, showing that it was good at identifying URLs. Problems in accurately classifying defacement, phishing, and malware were shown by a more thorough analysis of recall, precision, and F1-score. The Logistic Regression (LR) model also performed worse than the Random Forest Classifier in terms of accuracy. Problems in accurately identifying cases of defacement, phishing, and malware were highlighted by the confusion matrix, highlighting the difficulty of recognising subtle patterns linked to various cyber dangers. Owing to its excellent performance in lowering false positive and negative rates, the Gradient Boost model achieved a respectable accuracy score. While both the Gradient Boost and ANN models were accurate, the ANN model was far better at spotting harmless cases. Combining DL and NLP techniques with relevant characteristics from data preprocessing and EDA offers potential ways for cybersecurity applications to detect and mitigate the dangers posed by bad URLs.

6.2 Recommendations

The research shows that DL and NLP methods work well for classifying harmful URLs. In this light, businesses should fund R&D projects that use DL and NLP techniques to enhance their capacity to detect threats. It is critical to frequently add fresh samples of malware to training data as attackers are always changing their strategies. That way, machine learning models can detect new dangers as they emerge since they are constantly updated. In addition, ensemble approaches like the Random Forest Classifier have demonstrated encouraging outcomes when it comes to correctly detecting dangerous URLs. As a whole, detection systems may be made more accurate and resilient by investigating ensemble approaches, which merge several models. The processing of data and EDA also reveal useful details about dangerous URLs, like their length, character assessment, uncommon subdomains, and safe HTTP protocols. Better recognition of threats may be achieved by integrating these elements into the identification and training procedures. In addition, new threats are always appearing in the cybersecurity landscape, so models need to be reviewed, updated, and adjusted frequently. To keep the security system effective and reliable, it is vital to continuously evaluate it and make improvements.

6.3 Limitations

Through the use of data analysis, cyber security principles, and DL/NLP techniques, the study has found several limitations and opportunities for further research in the field of dangerous URL identification. Among the limitations are the difficulty in determining which deep learning algorithm is optimal, the demand for a substantial quantity of data for training, and the ongoing update of the training dataset with fresh malware samples.

6.4 Directions for Future Research

Potential avenues for future studies to solve the limitation include creating more sophisticated DL algorithms for fraudulent URL identification, finding better ways to update training datasets in actual time, and finding solutions to the problem of data scarcity. Current models might also use some work to improve their accuracy in categorising various forms of cyber threats, including phishing, malware, and defacement. Maintaining the models' efficacy and reliability in practical cybersecurity applications requires constant review, modification, and enhancement.

References

Abad, S., Gholamy, H., & Aslani, M. (2023). Classification of malicious URLs using machine learning. Sensors, 23(18), 7760. https://doi.org/10.3390/s23187760

Abdul Haseeb-ur-rehman, R.,M., Azana Hafizah, M. A., Mohammad, K. H., Khairul Akram, Z. A., Namoun, A., Tufail, A., & Ki-Hyung, K. (2023). High-speed network DDoS attack detection: A survey. Sensors, 23(15), 6850. https://doi.org/10.3390/s23156850

Abdulghani, A. A., Al-Bayatti, A., Saif, M., Jabbar, W. A., & Rassem, T. H. (2023). A honeybee-inspired framework for a smart city free of internet scams. Sensors, 23(9), 4284. https://doi.org/10.3390/s23094284

Afzal, S., Asim, M., Javed, A. R., Beg, M. O., & Baker, T. (2021). URLdeepDetect: A Deep Learning Approach for Detecting Malicious URLs Using Semantic Vector Models. *Journal of Network and Systems Management*, *29*(3). https://doi.org/10.1007/s10922-021-09587-8

Ahmed, S., Zahoor, A. K., Syed, M. M., Latif, S., Aslam, S., Mujlid, H., . . . Najam, Z. (2023). Effective and efficient DDoS attack detection using deep learning algorithm, multi-layer perceptron. Future Internet, 15(2), 76. https://doi.org/10.3390/fi15020076

Alanazi, A., & Gumaei, A. (2023). A decision-fusion-based ensemble approach for malicious websites detection. Applied Sciences, 13(18), 10260. https://doi.org/10.3390/app131810260

Alaoui, R. L., & El, H. N. (2022). Deep learning for vulnerability and attack detection on web applications: A systematic literature review. Future Internet, 14(4), 118. https://doi.org/10.3390/fi14040118

Alhogail, A., & Alsabih, A. (2021). Applying Machine Learning and Natural Language Processing to Detect Phishing Email. *Computers & Security*, *110*, 102414. https://doi.org/10.1016/j.cose.2021.102414

Ali, R., Ali, A., Iqbal, F., Hussain, M., & Ullah, F. (2022). Deep Learning Methods for Malware and Intrusion Detection: A Systematic Literature Review. *Security and Communication Networks*, *2022*, 1–31. https://doi.org/10.1155/2022/2959222

Aljabri, M., Alhaidari, F., Mohammad, R. M. A., Rami Mustafa, A. S. M., Alhamed, D. H., Altamimi, H. S., & Sara Mhd, B. C. (2022). An assessment of lexical, network, and content-based features for detecting malicious URLs using machine learning and deep learning models. Computational Intelligence and Neuroscience : CIN, 2022. https://doi.org/10.1155/2022/3241216

Aljabri, M., Altamimi, H.S., Albelali, S.A., Maimunah, A.H., Alhuraib, H.T., Alotaibi, N.K., Alahmadi, A.A., Alhaidari, F., Mohammad, R.M.A. and Salah, K., 2022. Detecting malicious URLs using machine learning techniques: review and research directions. IEEE Access. https://ieeexplore.ieee.org/iel7/6287639/6514899/09950508.pdf

Alshingiti, Z., Alaqel, R., Al-Muhtadi, J., Qazi Emad, U. H., Saleem, K., & Muhammad, H. F. (2023). A deep learning-based phishing detection system using CNN, LSTM, and LSTM-CNN. Electronics, 12(1), 232. https://doi.org/10.3390/electronics12010232

Aslan, Ö., Aktuğ, S. S., Ozkan-Okay, M., Yilmaz, A. A., & Akin, E. (2023). A comprehensive review of cyber security vulnerabilities, threats, attacks, and solutions. Electronics, 12(6), 1333. https://doi.org/10.3390/electronics12061333

Asmaa, R. O., Taie, S., & Shaheen, M. E. (2023). From phishing behavior analysis and feature selection to enhance prediction rate in phishing detection. International Journal of Advanced Computer Science and Applications, 14(5). https://doi.org/10.14569/IJACSA.2023.01405107

Atif, A. W., Li, Q., Zaland, Z., Shah, M., Dadan, K. B., Hussain, A., . . . Baryalai, M. (2023). A unified learning approach for malicious domain name detection. Axioms, 12(5), 458. https://doi.org/10.3390/axioms12050458

Bederna, Z., & Szadeczky, T. (2019). Cyber Espionage through Botnets. *Security Journal*, *33*. https://doi.org/10.1057/s41284-019-00194-6

Benavides, E., Fuertes, W., Sanchez, S., Sanchez, M. (2020). Classification of Phishing Attack Solutions by Employing Deep Learning Techniques: A Systematic Literature Review. In A. Rocha & R. Pereira (Eds.), *Developments and Advances in Defense and Security*. *Smart Innovation, Systems and Technologies* (pp. 51-64) Springer, Singapore. https://doi.org/10.1007/978-981-13-9155-2_5

Berman, D., Buczak, A., Chavis, J., & Corbett, C. (2019). A Survey of Deep Learning Methods for Cyber Security. *Information*, *10*(4), 122. https://doi.org/10.3390/info10040122

Birthriya, S.K., Jain, A.K. (2021). Analysis for Malicious URLs Using Machine Learning and Deep Learning Approaches. In M. Dave, R. Garg, M. Dua & J. Hussien (Eds.), *Proceedings of the International Conference on Paradigms of Computing, Communication and Data Sciences* (pp. 797-807). Springer, Singapore. https://doi.org/10.1007/978-981-15-7533-4_63

Cai, T., Liao, S., Xin, Y., & Lennon Y.C. Chang. (2018). Characteristics of cybercrimes: evidence from Chinese judgment documents. *Police Practice and Research*, *19*(6), 582–595. https://doi.org/10.1080/15614263.2018.1507895

Catal, C., Giray, G., Tekinerdogan, B., Kumar, S., & Shukla, S. (2022). Applications of deep learning for phishing detection: a systematic literature review. *Knowledge and Information Systems*, *64*(6), 1457–1500. https://doi.org/10.1007/s10115-022-01672-x

Chen, S., Lang, B., Chen, Y., & Xie, C. (2023). Detection of algorithmically generated malicious domain names with feature fusion of meaningful word segmentation and N-gram sequences. Applied Sciences, 13(7), 4406. https://doi.org/10.3390/app13074406

Collis, J., & Hussey, R. (2014) *Business Research: A Practical Guide for Undergraduate and Postgraduate Students.* Palgrave Macmillan.

Coyac-Torres, J., Sidorov, G., Aguirre-Anaya, E., & Hernández-Oregón, G. (2023). Cyberattack detection in social network messages based on convolutional neural networks and NLP techniques. Machine Learning and Knowledge Extraction, 5(3), 1132. https://doi.org/10.3390/make5030058

Darwish, S. M., Farhan, D. A., & Elzoghabi, A. A. (2023). Building an effective classifier for phishing web pages detection: A quantum-inspired biomimetic paradigm suitable for big data analytics of cyber attacks. Biomimetics, 8(2), 197. https://doi.org/10.3390/biomimetics8020197

Dutta, A. K. (2021). Detecting phishing websites using machine learning technique. PLoS One, 16(10) https://doi.org/10.1371/journal.pone.0258361

Ejaz, A., Mian, A. N., & Manzoor, S. (2023). Life-long phishing attack detection using continual learning. Scientific Reports (Nature Publisher Group), 13(1), 11488. https://doi.org/10.1038/s41598-023-37552-9

Elsadig, M., Ashraf, O. I., Basheer, S., Manal, A. A., Alshunaifi, S., Alqahtani, H., . . . Nagmeldin, W. (2022). Intelligent deep machine learning cyber phishing URL detection based on BERT features extraction. Electronics, 11(22), 3647. https://doi.org/10.3390/electronics11223647

Gómez, A., & Muñoz, A. (2023) 'Deep learning-based attack detection and classification in android devices', *Electronics*, 12(15), pp. 32-53.

Hajaj, C., Hason, N., & Dvir, A. (2022). Less is more: Robust and novel features for malicious domain detection. Electronics, 11(6), 969. https://doi.org/10.3390/electronics11060969

Hummel, P., Braun, M., Tretter, M., & Dabrock, P. (2021). Data sovereignty: A review. *Big Data & Society, 8*(1), 205395172098201. https://doi.org/10.1177/2053951720982012

Janet, B. and Kumar, R.J.A., 2021, March. Malicious URL detection: a comparative study. In 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS) (pp. 1147-1151).

IEEE. https://www.researchgate.net/profile/Shantanu-Maheshwari-2/publication/350931304_Malicious_URL_Detection_A_Comparative_Study/links/60da25cfa6fdccb745f09580/Malicious-URL-Detection-A-Comparative-Study.pdf

Johnson, C., Khadka, B., Basnet, R.B. and Doleck, T., 2020. Towards Detecting and Classifying Malicious URLs Using Deep Learning. J. Wirel. Mob. Networks Ubiquitous Comput. Dependable Appl., 11(4), pp.31-48. https://www.academia.edu/download/86678796/jowua-v11n4-3.pdf

Kabir, M. F., & Hartmann, S. (2018). Cyber security challenges: An efficient intrusion detection system design. *2018 International Young Engineers Forum (YEF-ECE)*. https://doi.org/10.1109/yef-ece.2018.8368933

Khonji, M., Iraqi, Y., & Jones, A. (2014). Phishing Detection: A Literature Survey. *IEEE Communications Surveys & Tutorials*, *15*(4), 2091–2121. https://doi.org/10.1109/surv.2013.032213.00009

Kim, H., & Lee, E. A. (2017). Authentication and Authorization for the Internet of Things. *IT Professional*, *19*(5), 27–33. https://doi.org/10.1109/mitp.2017.3680960

Kumar, N., Goel, V., Ranjan, R., Altuwairiqi, M., Alyami, H., & Simon, A. A. (2023). A blockchain-oriented framework for cloud-assisted system to countermeasure phishing for establishing secure smart city. Security and Communication Networks, 2023. https://doi.org/10.1155/2023/8168075

Maci, A., Santorsola, A., Coscia, A., & Iannacone, A. (2023). Unbalanced web phishing classification through deep reinforcement learning. Computers, 12(6), 118. https://doi.org/10.3390/computers12060118

Mahdavifar, S., & Ghorbani, A. A. (2019). Application of Deep Learning to Cybersecurity: A Survey. *Neurocomputing*, *347*. https://doi.org/10.1016/j.neucom.2019.02.056

Merlino, J. C., Asiri, M., & Saxena, N. (2022). DDoS cyber-incident detection in smart grids. Sustainability, 14(5), 2730. https://doi.org/10.3390/su14052730

Mohammed, M. A., Alornyo, S., Asante, M., & Essah, B. O. (2022). Intelligent detection technique for malicious websites based on deep neural network classifier. International Journal of Education and Management Engineering, 12(6), 45. https://doi.org/10.5815/ijeme.2022.06.05

Nagy, N., Aljabri, M., Shaahid, A., Amnah, A. A., Alnasser, F., Almakramy, L., . . . Alfaddagh, S. (2023). Phishing URLs detection using sequential and parallel ML techniques: Comparative analysis. Sensors, 23(7), 3467. https://doi.org/10.3390/s23073467

Natekin, A., & Knoll, A. (2013) ‘Gradient Boosting Machines, A Tutorial’, *Frontiers in Neurorobotics*, 7(21), pp. 1-15.

Ndichu, S., Kim, S., Ozawa, S., Ban, T., Takahashi, T., & Inoue, D. (2022). Detecting web-based attacks with SHAP and tree ensemble machine learning methods. Applied Sciences, 12(1), 60. https://doi.org/10.3390/app12010060

Olaniyan, R., Rakshit, S., & Vajjhala, N.R. (2023). Application of User and Entity Behavioral Analytics (UEBA) in the Detection of Cyber Threats and Vulnerabilities Management. In P. Chatterjee, D. Pamucar, M. Yazdani & D. Panchal (Eds.). *Computational Intelligence for Engineering and Management Applications* (pp. 419-426). Springer, Singapore. https://doi.org/10.1007/978-981-19-8493-8_32

Peng, T., Harris, I., & Sawa, Y. (2018). *Detecting Phishing Attacks Using Natural Language Processing and Machine Learning*. IEEE Xplore. https://doi.org/10.1109/ICSC.2018.00056

Prasad, S. D. V., & Rao, K. R. (2021). A novel framework for malicious URL detection using a hybrid model. Turkish Journal of Computer and Mathematics Education, 12(7), 68-76. Retrieved from https://www.proquest.com/scholarly-journals/novel-framework-malicious-url-detection-using/docview/2623612044/se-2

Rasheed, B., Khan, A., Ahsan Kazmi, ,S.M., Hussain, R., Md, J. P., & Doug, Y. S. (2021). Adversarial attacks on featureless deep learning malicious URLs detection. Computers, Materials, & Continua, 68(1), 921-939. https://doi.org/10.32604/cmc.2021.015452

Rea-Guaman, A.M., San Feliu, T., Calvo-Manzano, J.A., & Sanchez-Garcia, I.D. (2017). Comparative Study of Cybersecurity Capability Maturity Models. In A. Mas, A. Mesquida, R. O'Connor, T. Rout & A. Dorling (Eds.)s Software Process Improvement and Capability Determination. SPICE 2017. Communications in Computer and Information Science (100-113). Springer, Cham. https://doi.org/10.1007/978-3-319-67383-7_8

Rozi, M. F., Ozawa, S., Ban, T., Kim, S., Takahashi, T., & Inoue, D. (2022). Understanding the influence of AST-JS for improving malicious webpage detection. Applied Sciences, 12(24), 12916. https://doi.org/10.3390/app122412916

Sahingoz, O. K., Buber, E., Demir, O., & Diri, B. (2019). Machine learning based phishing detection from URLs. *Expert Systems with Applications*, *117*, 345–357. https://doi.org/10.1016/j.eswa.2018.09.029

Saleem R. A, Sundaravadivazhagan, B., Gamesman, P., Rajasekaran, J., & Karthikeyan, R. (2023). Weighted ensemble classifier for malicious link detection using natural language processing. *International Journal of Pervasive Computing and Communications*. https://doi.org/10.1108/ijpcc-09-2022-0312

Saleem Raja, A. S., Balasubaramanian, S., Al-Kaabi, A., Sharma, B., Chowdhury, S., Mehbodniya, A., . . . Bostani, A. (2023). Analysis of the performance impact of fine-tuned machine learning model for phishing URL detection. Electronics, 12(7), 1642. https://doi.org/10.3390/electronics12071642

Salloum, S., Gaber, T., Vadera, S., & Shaalan, K. (2021). Phishing Email Detection Using Natural Language Processing Techniques: A Literature Survey. *Procedia Computer Science*, *189*, 19–28. https://doi.org/10.1016/j.procs.2021.05.077

Samad, A., G Pradeepa, S. Mahalakshmi, & Jayakumar, M. S. (2023). Natural language based malicious domain detection using machine learning and deep learning. *Naučno-Tehničeskij*

Vestnik Informacionnyh Tehnologij, Mehaniki I Optiki, *23*(2), 304–312. https://doi.org/10.17586/2226-1494-2023-23-2-304-312

Sameen, M., Han, K. and Hwang, S.O., 2020. PhishHaven—An efficient real-time AI phishing URLs detection system. IEEE Access, 8, pp.83425-83443. https://ieeexplore.ieee.org/iel7/6287639/6514899/09082616.pdf

Sam-Shin, S., Seung-Goo Ji, & Sung-Sam, H. (2022). A heterogeneous machine learning ensemble framework for malicious webpage detection. Applied Sciences, 12(23), 12070. https://doi.org/10.3390/app122312070

Slayton, R. (2017). What Is the Cyber Offense-Defense Balance? Conceptions, Causes, and Assessment. *International Security*, *41*(3), 72–109. https://doi.org/10.1162/isec_a_00267

Somesha, M., Pais, A. R., Rao, R. S., & Rathour, V. S. (2020). Efficient deep learning techniques for the detection of phishing websites. *Sādhanā*, *45*(1). https://doi.org/10.1007/s12046-020-01392-4

Sun, L. (2021) 'Feature Engineering Framework based on Secure Multi-Party Computation in Federated Learning', *IEEE 23rd International Conference on High Performance Computing & Communications*, pp. 487-494.

Tsung, E. W. K. (2016) *The Philosophy of Management Research*. Routledge.

Uçar, E., Ucar, M., & İncetaş, M. O. (2019). A Deep learning approach for detection of malicious URLs. In *6th International Management Information Systems Conference* (pp. 12-20).

Retrieved from https://www.researchgate.net /publication/338477987_A_DEEP_LEARNING_APPROACH_FOR_DETECTION_OF_MALICIOUS_URLS

Umer, M., Sadiq, S., Karamti, H., Alhebshi, R. M., Alnowaiser, K., Ala’, A. E., . . . Ashraf, I. (2022). Deep learning-based intrusion detection methods in cyber-physical systems: Challenges and future trends. Electronics, 11(20), 3326. https://doi.org/10.3390/electronics11203326

Wei, X., Wei, X., Kong, X., Lu, S., Xing, W., & Lu, W. (2020). FMixCutMatch for semi-supervised deep learning. *Neural Networks*, *133*, 166–176. https://doi.org/10.1016/j.neunet.2020.10.018

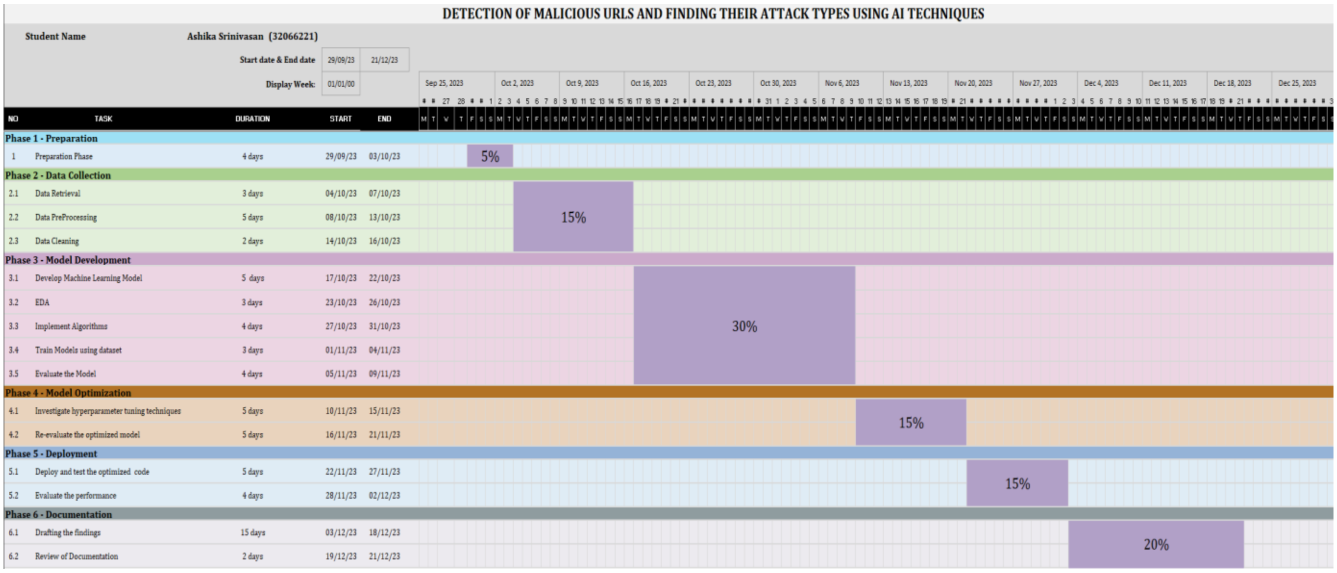
Wilson, J. (2010) *Essentials of Business Research: A Guide to Doing Your Research Project.* SAGE Publications.

Xue, D., Chi, Y., Wu, B., & Zhao, L. (2023). APT attack detection scheme based on CK sketch and DNS traffic. Sensors, 23(4), 2217. https://doi.org/10.3390/s23042217

Zhai, Y., Yang, L., Yang, J., He, L., & Li, Z. (2023). BadDGA: Backdoor attack on LSTM-based domain generation algorithm detector. Electronics, 12(3), 736. https://doi.org/10.3390/electronics12030736

Appendices

Appendix A - Research Project Plan



DETECTION OF MALICIOUS URLS AND FINDING THEIR ATTACK TYPES USING AI TECHNIQUES

Student Name Ashika Srinivasan (32066221)

Start date & End date 29/09/23 21/12/23

Display Week: 01/01/00

NO	TASK	DURATION	START	END
Phase 1 - Preparation				
1	Preparation Phase	4 days	29/09/23	03/10/23
Phase 2 - Data Collection				
2.1	Data Retrieval	3 days	04/10/23	07/10/23
2.2	Data PreProcessing	5 days	08/10/23	13/10/23
2.3	Data Cleaning	2 days	14/10/23	16/10/23
Phase 3 - Model Development				
3.1	Develop Machine Learning Model	5 days	17/10/23	22/10/23
3.2	EDA	3 days	23/10/23	26/10/23
3.3	Implement Algorithms	4 days	27/10/23	31/10/23
3.4	Train Models using dataset	3 days	01/11/23	04/11/23
3.5	Evaluate the Model	4 days	05/11/23	09/11/23
Phase 4 - Model Optimization				
4.1	Investigate hyperparameter tuning techniques	5 days	10/11/23	15/11/23
4.2	Re-evaluate the optimized model	5 days	16/11/23	21/11/23
Phase 5 - Deployment				
5.1	Deploy and test the optimized code	5 days	22/11/23	27/11/23
5.2	Evaluate the performance	4 days	28/11/23	02/12/23
Phase 6 - Documentation				
6.1	Drafting the findings	15 days	03/12/23	18/12/23
6.2	Review of Documentation	2 days	19/12/23	21/12/23

Project plan for detection of malicious activities

Appendix B - Ethics Form and Publication Procedure form

ETHICS FORM

Sheffield Hallam University

UREC2 RESEARCH ETHICS PROFORMA FOR STUDENTS UNDERTAKING LOW RISK PROJECTS WITH HUMAN PARTICIPANTS

This form is designed to help students and their supervisors to complete an ethical scrutiny of proposed research. The University Research Ethics Policy (www.shu.ac.uk/research/excellence/ethics-and-integrity/policies) should be consulted before completing this form. The initial questions are there to check that completion of the UREC 2 is appropriate for this study. The final responsibility for ensuring that ethical research practices are followed rests with the supervisor for student research.

Note that students and staff are responsible for making suitable arrangements to ensure compliance with the General Data Protection Act (GDPR). This involves informing participants about the legal basis for the research, including a link to the University research data privacy statement and providing details of who to complain to if participants have issues about how their data was handled or how they were treated (full details in module handbooks). In addition, the act requires data to be kept securely and the identity of participants to be anonymised. They are also responsible for following SHU guidelines about data encryption and research data management. Guidance can be found on the SHU Ethics Website www.shu.ac.uk/research/excellence/ethics-and-integrity

Please note that it is mandatory for all students to only store data on their allotted networked F drive space and not on individual hard drives or memory sticks etc.

The present form also enables the University and College to keep a record confirming that research conducted has been subjected to ethical scrutiny.

The form must be completed by the student and the supervisor and independently reviewed by a second reviewer or module leader (additional guidance can be obtained from your College Research Ethics Chair[1]). In all cases, it should be counter-signed and kept as a record showing that ethical scrutiny has occurred. Some courses may require additional scrutiny. Students should retain a copy for inclusion in their research project, and a copy should be uploaded to the relevant module Blackboard site.

Please note that it may be necessary to conduct a health and safety risk assessment for the proposed research. Further information can be obtained from the University's Health and Safety Website https://sheffieldhallam.sharepoint.com/sites/3069/SitePages/Risk-Assessment.aspx

SECTION A

1. Checklist questions to ensure that this is the correct form:

Health Related Research within the NHS, or His Majesty's Prison and Probation Service (HMPPS), or with participants unable to provide informed consent check list.

Question	Yes/No
Does the research involve? • Patients recruited because of their past or present use of the NHS	**No**
• Relatives/carers of patients recruited because of their past or present use of the	**No**

[1] College of Social Sciences and Arts - Dr. Antonia Ypsilanti (a.ypsilanti@shu.ac.uk)
College of Business, Technology and Engineering - Dr. Tony Lynn (t.lynn@shu.ac.uk)
College of Health, Wellbeing and Life Sciences - Dr. Nikki Jordan-Mahy (n.jordan-mahy@shu.ac.uk)

Question	Yes/No
NHS	
- Access to NHS staff, premises, or resources	**No**
- Access to data, organs, or other bodily material of past or present NHS patients	**No**
- Foetal material and IVF involving NHS patients	**No**
- The recently dead in NHS premises	**No**
- Prisoners or others within the criminal justice system recruited for health-related research	**No**
- Police, court officials, prisoners, or others within the criminal justice system	**No**
- Participants who are unable to provide informed consent due to their incapacity even if the project is not health related	**No**
- Is this an NHS research project, service evaluation or audit? *For NHS definitions please see the following website* http://www.hra.nhs.uk/documents/2013/09/defining-research.pdf	**No**

If you have answered **YES** to any of the above questions, then you **MUST consult with your supervisor** to obtain research ethics from the appropriate institution outside the university. This could be from the NHS or Her Majesty's Prison and Probation Service (HMPPS) under their independent Research Governance schemes. Further information is provided below. https://www.myresearchproject.org.uk/

2. Checks for Research with Human Participants

Question	Yes/No
1. Will any of the participants be vulnerable? *Note: Vulnerable people include children and young people, people with learning disabilities, people who may be limited by age or sickness, pregnancy, people researched because of a condition they have, etc. See full definition on ethics website in the document **Code of Practice for Researchers Working with Vulnerable Populations** (under the Supplementary University Polices and Good Research Practice Guidance)*	**No**
2. Are drugs, placebos, or other substances (e.g., food substances, vitamins) to be administered to the study participants or will the study involve invasive, intrusive, or potentially harmful procedures of any kind?	**No**
3. Will tissue samples (including blood) be obtained from participants?	**No**
4. Is pain or more than mild discomfort likely to result from the study?	**No**
5. Will the study involve prolonged or repetitive testing?	

Question	Yes/No
6. Is there any reasonable and foreseeable risk of physical or emotional harm to any of the participants? *Note: Harm may be caused by distressing or intrusive interview questions, uncomfortable procedures involving the participant, invasion of privacy, topics relating to highly personal information, topics relating to illegal activity, or topics that are anxiety provoking, etc.*	**No**
7. Will anyone be taking part without giving their informed consent?	**No**
8. Is the research covert? *Note: 'Covert research' refers to research that is conducted without the knowledge of participants.*	**No**
9. Will the research output allow identification of any individual who has not given their express consent to be identified?	**No**

If you have answered **YES** to any of these questions you are **REQUIRED** to complete and submit a UREC3 or UREC4 form. Your supervisor will advise. If you have answered **NO** to all these questions, then proceed with this form (UREC2).

3. General Project Details

Details	
Name of student	Ashika Srinivasan
SHU email address	C2066221@my.shu.ac.uk
Department/College	Masters in Artificial Intelligence/Sheffield Hallam University
Name of supervisor	Caren Fernandes
Supervisor's email address	caren.fernandes@shu.ac.uk
Title of proposed research	Detection of malicious URLs and finding their attack types using AI techniques
Proposed start date	29/09/2023
Proposed end date	21/12/2023
Background to the study and the rationale (reasons) for undertaking the research (500 words)	The Internet has become an essential aspect of modern life, facilitating everything from instantaneous global communication and data storage to online shopping and financial transactions. Unfortunately,

Details	
	the rise of harmful Uniform Resource Locators (URLs) and the sophisticated assaults they allow is another unintended consequence of this digital revolution. As per the view of Abad et al. (2023), the history and development of malicious URLs, the dangers they represent, and the countermeasures developed to prevent them are all covered in this section's context. Since its creation, the internet has gone a long way, expanding from a means of scholarly and military communication to a worldwide network that influences almost every area of contemporary life. As the Internet has developed, so have the strategies used by cybercriminals. The first phishing attempts appeared in the early days of the World Wide Web, marking the beginning of malicious URLs. As the reach of the internet increased to include more people, companies, and gadgets, so did the potential for abuse. Phishing, virus distribution, identity theft, and data breaches are just some of the cyber assaults that have used malicious URLs as a delivery mechanism. The widespread use of harmful URLs poses serious risks to both people and businesses. Cybercriminals are becoming more crafty in their attempts to trick visitors into accessing malicious websites. As cited by Alanazi & Gumaei (2023), phishing attacks, for example, often use apparently genuine URLs to lure victims into disclosing personal

Details	
	information like login credentials or financials. However, malware dissemination takes advantage of people's trusting nature by tricking them into visiting harmful websites using supposedly safe links. The use of malicious URLs is therefore not limited to the criminal community. Such URLs have been used by state-sponsored attackers to compromise government and infrastructure. There have been instances when these assaults have resulted in significant downtime, spying, and monetary losses. The magnitude of these dangers highlights the paramount need for solid, efficient defences. There has been an increase in the prevalence of malicious URLs, prompting the development of several tools for identifying and blocking them. As mentioned by Aljabri et al. (2022), signature-based and behaviour-based techniques are the two main buckets into which these strategies may be placed. In order to identify and block harmful URLs, signature-based detection uses patterns or signatures that have already been identified. However, this approach is not perfect since it cannot identify "zero-day" assaults that have never been observed before. In addition, malicious actors often adjust their methods in an effort to fool signature-based defences. Instead of relying on static signatures, behaviour-based detection

Details	
	examines how URLs really behave, including how they interact with people and what kinds of material they host. The advancement of behaviour-based detection systems has been greatly aided by machine learning techniques. As illustrated by Atif et al. (2023), because of their capacity to learn and adapt from massive datasets, these algorithms may identify previously unseen harmful URLs by analysing their behaviour and attributes.
Aims & research question(s)	This study aims to provide efficient techniques and tools to improve cybersecurity for detecting bad URLs and classifying the precise attack types they use. **Research Questions** 1. How can persuading techniques and methods be used to detect bad URLs while also recognising and separating the various attacks they could entail?
Methods to be used for: 1. Recruitment of participants 2. Data collection	**1.** No Participants will be involved in the data collection as it is collected from the external dataset detailed below. However, this

Details	
3. Data analysis	project involves a potential beneficiary who will participate in evaluating the final deliverable of this project. **2. Data Collection** Data Collection for this investigation will be a crucial stage, as the triumph of the suggested identification system and assault categorization model relies on the accessibility of an all-encompassing and varied collection of web addresses. To guarantee the effectiveness and significance of the analysis, a diverse approach will be embraced to collect both authentic and malevolent URL samples. The dataset is ***"https://www.kaggle.com/datasets/sid321axn/malicious-urls-dataset"*** which is acquired from Kaggle. Numerous data sources will be employed to encompass a broad array of real-life situations. This will involve performing web crawls from authentic websites to gather URLs that depict customary user conduct. Furthermore, trustworthy URL repositories and cybersecurity databases with recognised malicious URLs will be utilised to integrate a range of malevolent URL classifications. As per the view of Marcantoni et al. (2019), by merging information from these varied

Details	
	sources, the resulting dataset will accurately reflect the intricate and ever-changing characteristics of harmful URLs in their natural habitat. **3.Data Analysis** The gathered dataset of URLs will experience extensive data cleansing and pre-processing to guarantee its excellence and uniformity. Replicated records will be eliminated, and any absent data will be managed accordingly. The URLs will be standardised to consider differences in formats and guarantee consistency in the dataset. Preprocessing will additionally entail transforming textual data into an appropriate numerical format for analysis. By performing thorough data cleansing and preprocessing, the dataset will be purified and prepared for subsequent analysis. Exploratory data analysis will be conducted to acquire valuable understanding into the traits and arrangement of the URLs in the dataset. Explanatory figures and data visualisations will be utilised to investigate the spread of authentic and malevolent URLs, URL sizes, domain durations, and other pertinent characteristics. As illustrated by Li et al. (2020), this investigative process will offer a more profound comprehension of the dataset, unveiling possible patterns and

Details	
	connections that might impact the model's effectiveness. A varied mixture of machine learning models will be utilised to create the identification system and assault categorization model. Various algorithms, such as logistic regression, assist vector machines, chance forests, gradient boosting, and neural networks, will be investigated and contrasted for their effectiveness. Collective approaches can also be employed to further improve the precision and resilience of the model. As cited by Liu et al. (2021), the chosen models will be educated on the training dataset, utilising the characteristics extracted during the preprocessing stage.
Outline the nature of the data held, details of anonymisation, storage and disposal procedures as required.	The data is acquired from "https://www.kaggle.com/datasets/sid321axn/malicious-urls-dataset" from the Kaggle. The analysis will be conducted using the Python programming language, resulting in the generation of Python file with the appropriate extension. The licence is attached below.

Details	
	 CC0 1.0 DEED CC0 1.0 Universal See the legal code No Copyright [illegible] Other Information [illegible] Other Rights [illegible] Learn more. Endorsement [illegible] Learn more. Citation [illegible] Learn more Who is the affirmer? [illegible] Notice [illegible]

4. Research in External Organisations

Question	Yes/No
1. Will the research involve working with/within an external organisation (e.g., school, business, charity, museum, government department, international agency, etc.)?	**No**
2. If you answered YES to question 1, do you have granted access to conduct the research from the external organisation? *If YES, students please show evidence to your supervisor. You should retain this evidence safely.*	**No**
3. If you do not have permission for access is this because: A. you have not yet asked B. you have asked and not yet received an answer C. you have asked and been refused access *Note: You will only be able to start the research when you have been granted access.*	**No**

5.Research with Products and Artefacts

Question	Yes/No
1. Will the research involve working with copyrighted documents, films, broadcasts, photographs, artworks, designs, products, programs, databases, networks, processes, existing datasets, or secure data?	**Yes**
2. If you answered YES to question 1, are the materials you intend to use in the public domain? *Notes: 'In the public domain' does not mean the same thing as 'publicly accessible'.* • *Information which is 'in the public domain' is no longer protected by copyright (i.e., copyright has either expired or been waived) and can be used without permission.* • *Information which is 'publicly accessible' (e.g., TV broadcasts, websites, artworks, newspapers) is available for anyone to consult/view. It is still protected by copyright even if there is no copyright notice. In UK law, copyright protection is automatic and does not require a copyright statement, although it is always good practice to provide one. It is necessary to check the terms and conditions of use to find out exactly how the material may be reused etc.* *If you answered YES to question 1, be aware that you may need to consider other ethics codes. For example, when conducting Internet research, consult the code of the Association of Internet Researchers; for educational research, consult the Code of Ethics of the British Educational Research Association.*	**Yes**
3. If you answered NO to question 2, do you have explicit permission to use these materials as data? *If YES, please show evidence to your supervisor.*	**NA**
4. If you answered NO to question 3, is it because: A. you have not yet asked permission B. you have asked and not yet received and answer C. you have asked and been refused access. *Note: You will only be able to start the research when you have been granted permission to use the specified material.*	**A/B/C**

SECTION B

HEALTH AND SAFETY RISK ASSESSMENT FOR THE RESEARCHER

1. **Does this research project require a health and safety risk assessment for the procedures to be used?** (Discuss this with your supervisor)

   □ Yes
   ✓ No

   If **YES** the completed Health and Safety Risk Assessment form should be attached. A standard risk assessment form can be generated through the Awaken system (https://shu.awaken-be.com). Alternatively if you require more specific risk assessment, e.g. a COSHH, attach that instead.

2. **Will the data be collected fully online (no face-to-face contact with participants)?**

   ✓ Yes (See the safety guidance for online research[2] and **go to question 7b**)
   □ No (Go to question 3)

3. **Will the proposed data collection take place on campus?**

   □ Yes (Please answer questions 5 to 8)
   ✓ No (Please complete all questions and consult with your supervisor))

4. **Where will the data collection take place?**
   (Tick as many as apply if data collection will take place in multiple venues)

	Location	**Please specify**
□	Researcher's Residence	
□	Participant's Residence	

[2] Safety guidance for online research includes information on how to set up online surveys and/or conduct online interviews/focus groups. These guidelines can be found in BB. Please check with your supervisor/module leader.

	Location	**Please specify**
□	Education Establishment	
□	Other e.g., business/voluntary organisation, public venue	
□	Outside UK	

5. How will you travel to and from the data collection venue?

□ On foot □ By car □ Public Transport
□ Other (Please specify)

Please outline how you will ensure your personal safety when travelling to and from the data collection venue.

6. How will you ensure your own personal safety whilst at the research venue?

7. Are there any potential risks to your health and wellbeing associated with either (a) the venue where the research will take place and/or (b) the research topic itself?

✓ None that I am aware of
□ Yes (Please outline below including steps taken to minimise risk)

8. If you are carrying out research off-campus, you must ensure that each time you go out to collect data you ensure that someone you trust knows where you are going (without breaching the confidentiality of

your participants), how you are getting there (preferably including your travel route), when you expect to get back, and what to do should you not return at the specified time.

Please outline here the procedure you propose using to do this.

Insurance Check

The University’s standard insurance cover will not automatically cover research involving any of the following:

i) Participants under 5 years old
ii) Pregnant women
iii) 5000 or more participants
iv) Research being conducted in an overseas country
v) Research involving aircraft and offshore oil rigs
vi) Nuclear research
vii) Any trials/medical research into Covid 19

If your proposals do involve any of the above, please contact the Insurance Manager directly (fin-insurancequeries-mb@exchange.shu.ac.uk) to discuss this element of your project.

Adherence to SHU Policy and Procedures

Ethics sign-off	
Personal statement	
I can confirm that: • I have read the Sheffield Hallam University Research Ethics Policy and Procedures • I agree to abide by its principles.	
Student	
Name: Ashika Srinivasan	Date: 25-10-2023
Signature:	
Supervisor ethical sign-off	
I can confirm that completion of this form has not identified the need for ethical approval by the TPREC/CREC or an NHS, Social Care, or other external REC. The research will not commence until any approvals required under Sections 4 & 5 have been received and any necessary health and safety measures are in place.	

Ethics sign-off	
Name: Caren Fernandes	Date:17/11/2023
Signature:	
Independent Reviewer ethical sign off	
Name:	Date:
Signature:	

Please ensure that you have attached all relevant documents. Your supervisor must approve them before you start data collection:

Documents	**Yes**	**No**	**N/A**
Research proposal if prepared previously	✓	□	□
Any recruitment materials (e.g., posters, letters, emails, etc.)	□	□	□
Participant information sheet[3]	✓	□	□
Participant consent form[4]	✓	□	□
Details of measures to be used (e.g., questionnaires, etc.)	✓	□	□
Outline interview schedule / focus group schedule	□	□	□
Debriefing materials	□	□	□
Health and Safety Risk Assessment Form	□	□	□

[3] It is mandatory to attach the Participant Information Sheet (PIS)

[4] It is mandatory to attach a Participant Consent Form, unless it is embedded in an online survey, in which case your supervisor must approve it before you start data collection.

PUBLICATION PROCEDURE FORM

Sheffield Hallam University | College of Business, Technology and Engineering

Research Skills and Dissertation Module (55-706556).

PUBLICATION PROCEDURE FORM

In this module, while you create your own research question or topic area, your supervisor makes a significant intellectual contribution to this work as the research progresses. Your supervisor will make the decision on whether your work merits publication based on the quality of the work you have produced. Your supervisor will co-author the paper for publication with you and your supervisor will both be listed as authors. You are required to sign the declaration below to confirm that you understand and will follow this procedure.

Declaration:

I **Ashika Srinivasan** confirm that I understand will comply with the Publication Procedure outlined in the Module Handbook and the Blackboard Site.

G5

Student:	Signature	Date 18/01/2024
Supervisor:	Signature	Date 18/01/2024

Appendix C - Participant Information Sheet, Consent form and Beneficiary feedback form

PARTICIPANT INFORMATION SHEET

Title of Project: Detection of malicious URLs and finding their attack types using AI techniques

Dear Jing,

I invite you to participate in a groundbreaking research study focused on the "Detection of Malicious URLs and Finding Their Attack Types Using AI Techniques." Your participation is crucial to advancing my understanding of cybersecurity threats and developing more effective strategies to combat malicious activities on the internet.

Your participation in this study will contribute significantly to the development of intelligent systems that can proactively identify and mitigate potential cybersecurity threats. Sharing your insights and experiences plays a vital role in advancing the field of cybersecurity and helping create a safer online environment for individuals and organizations.

The primary goal of this research is to leverage advanced Artificial Intelligence (AI) techniques to detect and analyse malicious Uniform Resource Locators (URLs) — web addresses designed to compromise security. By understanding the characteristics of these URLs, I aim to enhance the ability to identify and classify various attack types such as phishing, malware distribution, and defacement.

Participants in this study were selected based on specific criteria that align with the objectives of my research. I aimed to include individuals with diverse experiences in the AI field and perspectives related to online security and interactions with URLs.

Your participation in this research is completely voluntary. If at any point you wish to no longer take part in the research you have the right to withdraw at any time and there will be no pressure to stay.

The beneficiary feedback form questionnaire should take no longer than 10 minutes to complete fully.

You will have the opportunity to discuss your participation in the research once it has been conducted. You will be permitted to ask any more questions or seek clarification. All the information you give will be confidential and only used for this research.

All data collected during the study will be anonymized and handled with the utmost care.

Details of who to contact if you have any concerns or if adverse effects occur after the study are given below.

Researcher/ Research Team Details:

You should contact the Data Protection Officer if:	**You should contact the Head of Research Ethics (Dr Mayur Ranchordas) if:**
• you have a query about how your data is used by the University • you would like to report a data security breach (e.g. if you think your personal data has been lost or disclosed inappropriately) • you would like to complain about how the University has used your personal data DPO@shu.ac.uk	• you have concerns with how the research was undertaken or how you were treated ethicssupport@shu.ac.uk
Postal address: Sheffield Hallam University, Howard Street, Sheffield S1 1WBT Telephone: 0114 225 5555	

PARTICIPANT CONSENT FORM

TITLE OF RESEARCH STUDY:

Please answer the following questions by ticking the response that applies

		YES	NO
1.	I have read the Information Sheet for this study and have had details of the study explained to me.	☒	□
2.	My questions about the study have been answered to my satisfaction and I understand that I may ask further questions at any point.	☒	□
3.	I understand that I am free to withdraw from the study within the time limits outlined in the Information Sheet, without giving a reason for my withdrawal or to decline to answer any particular questions in the study without any consequences to my future treatment by the researcher.	☒	□
4.	I agree to provide information to the researchers under the conditions of confidentiality set out in the Information Sheet.	☒	□
5.	I wish to participate in the study under the conditions set out in the Information Sheet.	☒	□
6.	I consent to the information collected for the purposes of this research study, once anonymised (so that I cannot be identified), to be used for any other research purposes.	☒	□

Participant’s Signature: ____ Date 17/01/2024

Participant’s Name (Printed): _Jing Wang__________________________

Contact details: [redacted]

____

Researcher’s Name (Printed): ASHIKA SRINIVASAN

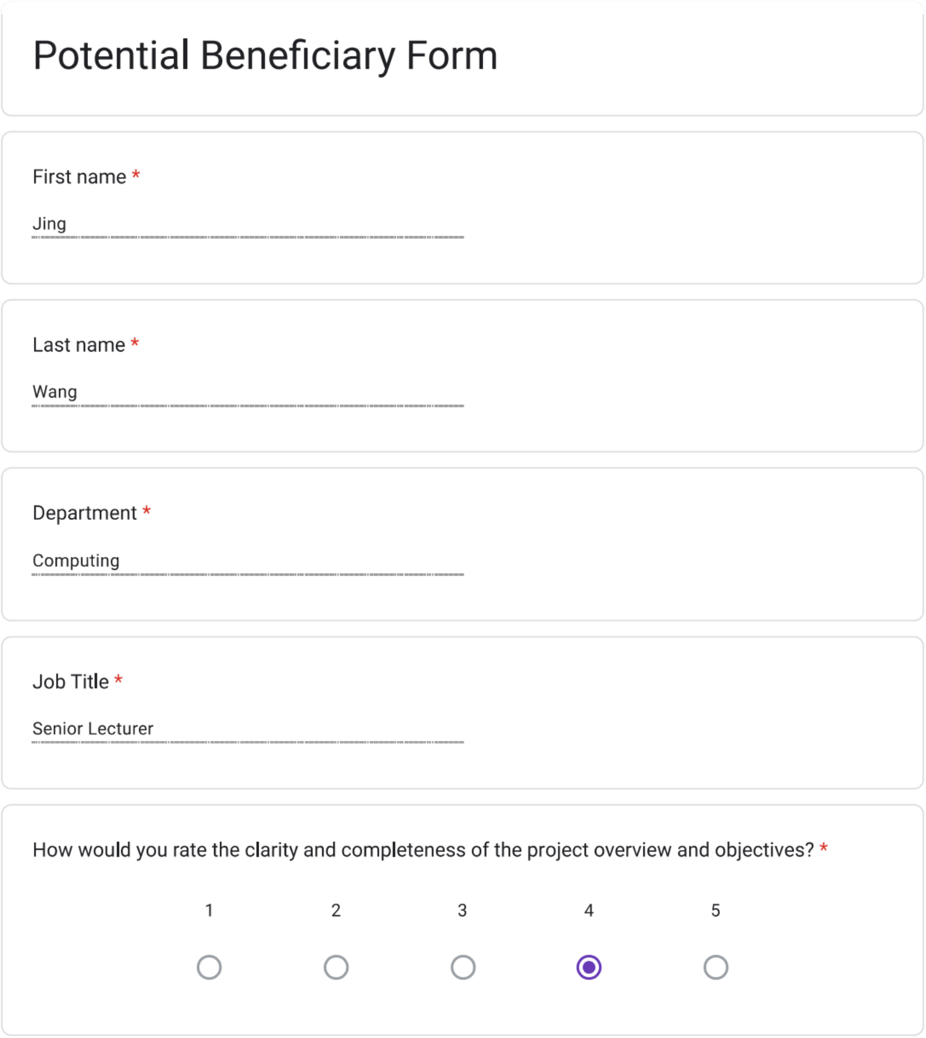
Researcher’s Signature:

Researcher's contact details: [redacted]

(Name, address, contact number of investigator)

BENEFICIARY FEEDBACK FORM

beneficiary_feedback.pdf



Potential Beneficiary Form

First name *

Jing

Last name *

Wang

Department *

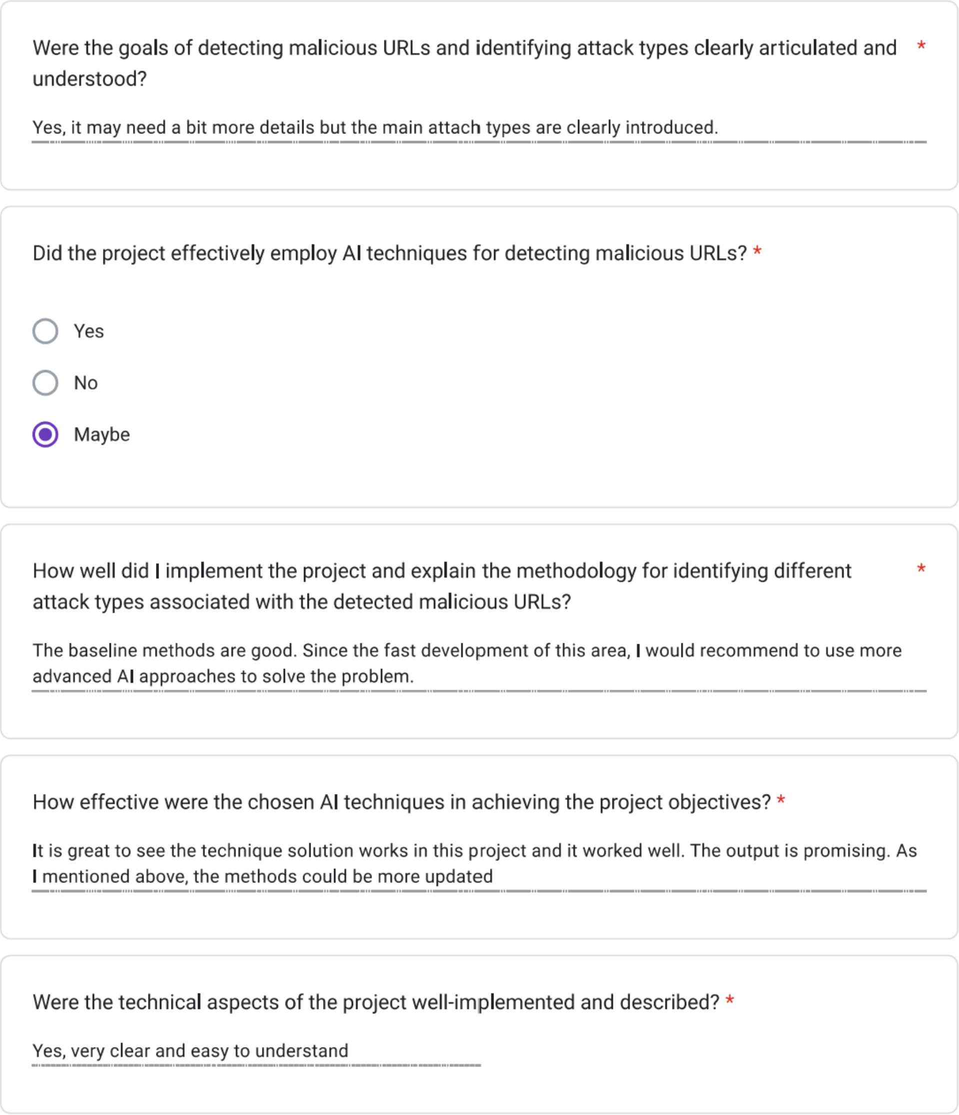
Computing

Job Title *

Senior Lecturer

How would you rate the clarity and completeness of the project overview and objectives? *

1	2	3	4	5
○	○	○	●	○



Were the goals of detecting malicious URLs and identifying attack types clearly articulated and understood? *

Yes, it may need a bit more details but the main attach types are clearly introduced.

Did the project effectively employ AI techniques for detecting malicious URLs? *

○ Yes

○ No

● Maybe

How well did I implement the project and explain the methodology for identifying different attack types associated with the detected malicious URLs? *

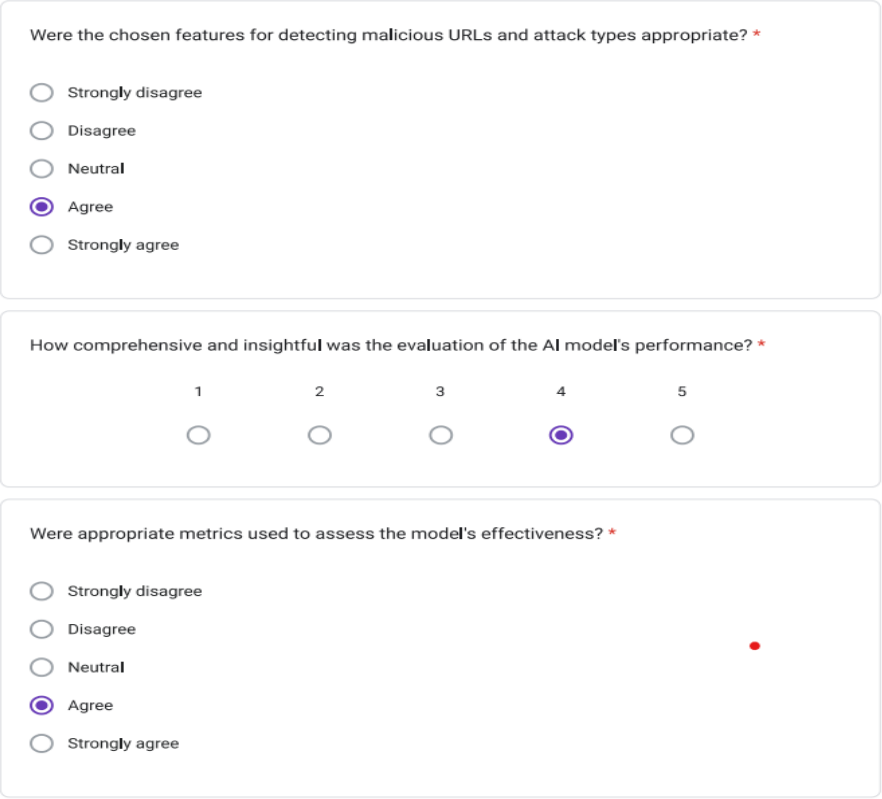
The baseline methods are good. Since the fast development of this area, I would recommend to use more advanced AI approaches to solve the problem.

How effective were the chosen AI techniques in achieving the project objectives? *

It is great to see the technique solution works in this project and it worked well. The output is promising. As I mentioned above, the methods could be more updated

Were the technical aspects of the project well-implemented and described? *

Yes, very clear and easy to understand



Were the chosen features for detecting malicious URLs and attack types appropriate? *

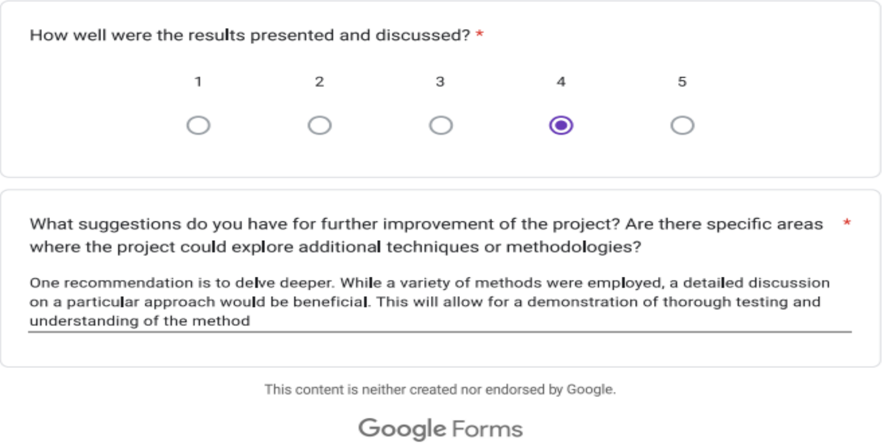
○ Strongly disagree
○ Disagree
○ Neutral
● Agree
○ Strongly agree

How comprehensive and insightful was the evaluation of the AI model's performance? *

1	2	3	4	5
○	○	○	●	○

Were appropriate metrics used to assess the model's effectiveness? *

○ Strongly disagree
○ Disagree
○ Neutral
● Agree
○ Strongly agree



How well were the results presented and discussed? *

1	2	3	4	5
○	○	○	●	○

What suggestions do you have for further improvement of the project? Are there specific areas where the project could explore additional techniques or methodologies? *

One recommendation is to delve deeper. While a variety of methods were employed, a detailed discussion on a particular approach would be beneficial. This will allow for a demonstration of thorough testing and understanding of the method

This content is neither created nor endorsed by Google.

Google Forms

Appendix D - Dataset Link

Here is the secondary dataset link, https://www.kaggle.com/datasets/sid321axn/malicious-urls-dataset

Appendix E - Software Code Link

Below is the link to the code

https://colab.research.google.com/drive/1kdSdYmzjg7yX56zfwesEPqK4tLaE9xoM?usp=sharing

Appendix F - Pre-Submission Checklist

PRE-SUBMISSION CHECKLIST

Before submitting the final version of your dissertation, please complete the checklist below.

Is the Cover Page as specified in the Dissertation guidelines?	✓
Have you included a confidentiality statement on the Title Page, if applicable?	✓
Have you included an Abstract?	✓
Have you included your Research Project Plan as an Appendix?	✓
Have you included the appropriate Research Ethics Proforma as an Appendix?	✓
Have you included your Publication Procedure form as an Appendix?	✓
Have you included a link to your secondary data as an Appendix?	✓
If applicable, have you included a link to your code as an Appendix?	✓