Detecting business email compromise and classifying for countermeasures

Pubudu Gayan Buddhika

HND, MSc-IT, PGDipBM, MISDF

A thesis submitted to the graduate faculty of design and creative technologies

Auckland University of Technology

In fulfilment of the

Requirement of the degree of

Master of Philosophy

School of Engineering, Computer and Mathematical Sciences

Auckland, New Zealand

2023

Declaration

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the qualification of any other degree or diploma of a University or other institution of higher learning, except where due acknowledgement is made in the acknowledgements.

Pubudu Gayan Buddhika

Acknowledgements

First, I would like to dedicate this thesis to all the health sector professionals who sacrificed their lives to recover the world from the COVID pandemic. Secondly, I would like to thank Auckland University of Technology for the opportunity given to me to improve my research knowledge skills. Additionally, I am eternally indebted to my lovely mother, who is behind me all the time and encourages me when I go through difficult times.

Furthermore, I would like to express my deepest gratitude to my supervisor Prof. Brian Cusack for all the guidance and support he has given me since the start date of this thesis. Significantly, his support during the COVID time was unforgettable when I struggled.

Finally, I would like to thank my angel (daughter), Oshdhi, who asked me many questions regarding my research and was patient with me.

Abstract

The use of email has evolved radically since 1965 when the first email was sent. It was just a simple text, like a paste note, that anyone could see who used that computer. Today emails are multimedia, have their own servers or cloud service, and are at the core of global communications. With the technology advancements and Internet expansion, email has become an essential tool for individuals and organisations as a communication tool. Today when millions of users use email to exchange messages and information, perpetrators are attracted to steal information from email messages. The commercial use of email has heightened the motivation of hackers and attackers to compromise email communications and to exploit them for their own gain. Therefore, IT security experts have introduced heavy encryption methods to protect the email message and its contents, plus various protocols and security standards have been implemented to protect email communication channels. However, Rose (2021) cites that most email and IT-related security issues occur today because of human errors. These human errors occur because of a lack of awareness of the security threats, failure to follow instructions, and insecure local devices.

In this research the first phase of experimental work is collecting the pilot spam emails to build the corpora for training the NLP model. Initially the NLP model uses binary classification to categorise the spam into harmful and harmless emails. In the second stage, I build the python program to classify spam and ham emails using NLP. At the same time, I set up the isolated mock network to collect the spam emails according to the user behaviour within the controlled and scripted actions. In the third phase, I train the NLP model and use the collected spam email to identify how accurate is the NLP model. These findings lead to accuracy, efficiency, and information on how user behaviour influences the number of incoming spam emails based on what users do on the Internet. It is concluded that security models can be tricked by the similarity of harmful and harmless emails. The hackers and attackers are highly skilled at crafting similarity. The trained NLP model enhances protection, adaptation to new threats, and learning, but sometimes it is still not enough to identify all spam emails. Therefore, the in the last phase of research I design a new framework (Figure 5.1) to improve business email security by identifying spam and other false emails before they come to the user inboxes. The research results also signal the importance of training and self-monitoring of all behaviours used on the Internet as this influences the amount and type of spam attacks received.

Table of Contents

Table of Figures

List of Tables

List of Abbreviations

ABC	Australian Broadcasting Corporation
AES	Advanced Encryption Standard
AI	Artificial Intelligence
API	Application Programming Interface
BEC	Business Email Compromise
CEO	Chief Executive Officer
COVID	Coronavirus disease
CSV	Comma separated values
DHA	Directory Harvest Attacks
DKIM	Domain Keys Identified Mail
DLL	Dynamic link library
DNA	Deoxyribonucleic Acid
FBI	Federal Bureau of Investigation
FN	False Negative
FP	False Positive
HTTP	Hypertext Transfer Protocol
IEEE	Institute of Electrical and Electronics Engineers
IMAP	Internet Messages Access Protocol
ISO	International Organization for Standardization
MAA	Message Access Agent
MILK	Machine Learning toolkit
ML	Machine Learning
MTA	Message Transfer Agent
MUA	Mail User Agent
NDR	Non delivery report
NIST	National Institute of Standards and Technology
NLG	Natural Language Generation
NLP	Natural Language Processsing
NLU	Natural Language Understanding

NN	Neural Network
POP	Post Office protocol
QR	Quick Response
RSA	Rivest-Shanir-Adleman
SMTP	Simple Mail Transfer Protocol
SSL	Secure Sockets Layer
SVM	Support vector machine
TLS	Transport Layer Security
TN	True Negative
TP	True Positive
URL	Uniform Resource Locator
VPN	Virtual Private Network

Chapter 1

Introduction

1.0 BACKGROUND

The use of email has evolved radically since 1965 when the first email was sent. It was just a simple text, like a paste note, that anyone could see who used that computer. Today emails are multimedia, have their own servers or cloud service, and are at the core of global communications. With the technology advancements and internet expansion, email has become an essential tool for individuals and organisations as a communication tool. Before the internet, email was sent via intranets and used terminal connectivity to establish the connection between a sender and the receiver. However, with the help of new technology, email messages also get new components such as an email inbox, message protocols, delivery notification, and so on. Ease of use and interoperability make the email system popular, and, generally, more than 500 million users are using emails daily (Fahad, 2020).

When millions of users use email to exchange messages and information, perpetrators are attracted to steal information from email messages. In early email use, there was little security enforcement or protocols to protect email communications as email messages were limited to the intranet. However, with commercialising the service, email users and service providers are more concerned about security (Five key points the public sector should consider when switching email service providers, 2023). Therefore, IT security experts tried to protect email services from hackers and cyber-criminals. To protect email and its communication channels, security experts introduced encryption methods to protect the email message and its contents, plus introduced various protocols and security standards to protect email communication channels. Email security enforcement not only protects email messages but closes an entry point to the corporate network for hackers to steal other valuable information or data stored within the corporate network. Spam emails, phishing, and malware delivery are the most common email attack methods that cybercriminals are employing today. To protect email inboxes from these false emails, IT security experts use various enforcements, such as virus guards, encryption, checkpoints, and VPNs. However, hackers also use various techniques and technologies to overcome these security barriers.

Rose (2021) cites that most email and IT-related security issues occur because of human errors. These human errors occur because of a lack of awareness of the security threats, failure to follow instructions, and not up-to-date patches on local devices. With the COVID -19 pandemic, almost all employees are working from home, and, the Australian Security & Investments Commission website has stated that scammers are taking advantage of this pandemic environment because of the pandemic (Scammers taking advantage of COVID-19 to target small businesses, 2021). Scammers use special techniques in this pandemic period to swindle business email users, such as selling vaccines, taking bank authorisation details, fake delivery orders for masks, inexplicable urgency, like suppliers requesting advanced payments because of the COVID pandemic and many more.

Most business organisations have recognised that educating their email users is one of the best ways to minimise email-related threats. The problem is that it is difficult to differentiate a genuine email and a spam email, even for experts. Often phishing emails and genuine emails look precisely the same in plain sight. Identifying the false emails before they go into legitimate users' inboxes is the best way to minimise business email compromise (BEC). However, this is currently an impossible task and research is ongoing to catch up with the scammers and hackers, and to improve email security enforcement. They also upskill their techniques and methods to outwit security provisions. Benishti (2022) stated in Forbes magazine that modern phishing and spam emails do not contain malicious URLs or payloads, but the attackers use social engineering techniques to encourage email users to provide valuable information. To overcome this challenge, researchers are investigating Machine Learning and Natural Language Processing techniques to identify fake emails from legitimate emails. Using NLP engines email security engines can identify the pattern of the message text arrangement, grammatical mistakes, and syntax mistakes. Plus, ML-based security solutions can identify potential future security attacks as well (Benishti, 2022).

In this research, the latest NLP techniques are to be investigated and tested against real spam emails. The literature review elaborates the anatomy of the business email system and the architecture to understand the email system attack vectors. Furthermore, email message essential components and the various types of protocols are defined. Consequently, Business Email Compromise (BEC) attacks can be identified, and the pattern types of spammers used to compromise business email

systems disclosed. Natural Language Processing (NLP) is at the cutting edge of current email security research. In chapter 2 it is defined and in chapter 3 the research question and hypothesis are outlined. Using the combination of research findings and previous research publications, this research thesis defines a research methodology for NLP, and delivers a proposed new email security framework using NLP pattern classification to reduce the risk of BEC. The title of "Detecting business email compromise and classifying for countermeasures." Chosen to represent the core of the research and the deliverable. The research question, "What patterns are detected in email attacks on business systems?", is selected to guide the research project and to focus on the potential AI NLP contribution to email security.

1.1 MOTIVATION

Section 1.0 has described the background of BEC and the scope of the proposed research. In this section, the motivation factors driving the researcher are outlined. The research topic is based on the following two motivations. The understanding of the patterns of behaviour for BEC, and how countermeasures can be implemented through an up-to-date framework using NLP AI.

Email is a leading communication channel in the business world today (Taylor , 2019). Most government organisations and internet transaction applications request an email address from their service receivers to keep the communication channel or to inform valuable information to the users. Such as policy changes, two factor authentication verification, payments, legal matters, and so on. With the availability of the internet, internet users and the corporate sector also use email to exchange various forms of valuable information. It contains information of commercial value or classified information from the public entities. Therefore, security is critical for email accounts, the transmission of emails, and the protection of email contents. However, Kee (2021) stated in Forbes magazine that cyber-attacks and email threats are real and a real threat to email communications. Furthermore, according to Lila (1990), the HTTP protocol is no longer accepted as a secure protocol, and the later improvement that added a secure signed certificate to enhance its security for HTTPS is necessary (Kee, 2021). Similarly, the email protocol SMTP (Simple Mail Transfer Protocol) moved to Secure Multipurpose Internet Mail Extensions (S/MIME). Nevertheless, email spammers and hackers also upskilled their techniques and technologies to bypass

these security improvements. According to the spanning security (2022) website, most email attacks are phishing or social engineering attacks now. Moreover, according to the spanning website, most of these phishing attacks happened because of human errors (Cyberattacks 2021: Phishing, Ransomware & Data Breach Statistics From the Last Year, 2022).

With the rapidly changing vectors for attack on email systems and the threats to business emails and related systems, new security countermeasures to protect the email system are required. Furthermore, Louis argues in his article that with the help of ML spam filters can understand the future threat models based on location, user behaviour patterns and device locations (Columbus, 2020). Most applications and smart devices collect user behaviour information and location details for marketing purposes and sell them to third parties. Therefore, these data can be used for spreading targeted spam emails and phishing attacks. With the traditional email security protection system, this cannot be achieved. Traditional email security systems do not have a trained model that can understand user behaviour, learn, and predict the evolving threat model.

The second motivating point is that AI and ML are continually acquiring improvements plus reducing resource requirements for algorithm computing. In the past, processing AI-related complex algorithms needed massive processing power and expensive equipment. Additionally, unlike old spam emails, the new ones do not contain any fake URLs to redirect users to malicious websites. Instead of fake URLs they using search engines URL redirect features. In Sophos security website they explained it with the example (How scammers abuse Google Search's open redirect feature, 2020). Furthermore attackers sends emails that contain QR codes or similar techniques to redirect users to the malicious websites. By using ML, it can analyse these kinds of images, the contents, email bodies, and the sender characteristics.

Another motivation is that analysing the social engineering attacks is quite interesting, especially with the user internet search footprint relationship to incoming spam. Nicholas (2022) stated that the US FBI crime report in 2021, reported more than 300,000 incidents related to social engineering attacks. However, these numbers are small, and the actual numbers can be far greater when unreported attacks are also counted. Hence, information security engineers need a robust framework that can recognise social engineering phishing email attacks

more than ever to protect their legitimate corporate email users. Considering all of the above factors and influences that motivated me to investigate this topic I am highly motivated to learn more, and to propose a current threats countermeasure framework for protecting business email accounts from various types of attacks, but in particular social engineering attacks.

1.2 RESEARCH APPROACH AND FINDINGS

The initial approach has been taken by reviewing literatures that related to BEC, NLP, and current spam filter technologies. This work is to establish a contemporary context for the study. Then I review four similar research topic reports to understand how other researchers have set up their methodologies, methods, and tools. The related research publications help to speed up, inform, and to focus the NLP experimental work. The evidence from the experiments will help to find the answer to the main research question. Cybercriminals and spam email senders are using various techniques to breach business email security, and research must be current. Additionally, BEC is an changing topic with the existing technology, and some past research publications may be outdated. Therefore, when selecting the literature for this research I selected the most recent publications as they addressed the most recent vulnerabilities and techniques of attack and defence.

In this research the first phase of experimental work is collecting the pilot spam emails to build the corpora for training the NLP model. Initially the NLP model uses binary classification to categorise the spam into harmful and harmless emails. In the second stage, I build the python program to classify spam and ham emails using NLP. At the same time, I set up the isolated mock network to collect the spam emails according to the user behaviour within the controlled and scripted actions. In the third phase, I train the NLP model and use the collected spam email to identify how accurate is the NLP model. These findings will lead to accuracy, efficiency, and information on how user behaviour influences the number of incoming spam emails based on what users do on the Internet.

This research shows that security models can be tricked by the similarity of harmful and harmless emails. The hackers and attackers are highly skilled at crafting similarity. The trained NLP model enhances protection, adaptation to new threats, and learning, but sometimes it is still not enough to identify all spam emails. Especially when spam emails contain greetings and names with the email

address recipient then the NLP based model is not enough to identify the incoming spam emails. Therefore, the last phase of research suggests a new framework to improve business email security by identifying spam and other fake emails before they come to the user inboxes. The research results also signal the importance of training and self-monitoring of all behaviours used on the Internet as this influences the amount and type of spam attacks received.

1.3 STRUCTURE OF THE THESIS

This section, describes the structure of the research thesis and an outline of each chapter.

The formalities are mandatory and set out the title, authorisation declarations, and tables for contents for reader access to the research. This chapter 1 provides a concise summary of the research and the findings.

In this research, the latest natural language processing (NLP) techniques are to be investigated and tested against real spam emails. The literature review elaborates the anatomy of the business email system and the architecture to understand the email system attack vectors. Furthermore, the email message components and the various types of protocols are defined. NLP is at the cutting edge of current email security research. In chapter 2 NLP is defined and in chapter 3 the research question and hypothesis are outlined. Using the combination of research findings and previous research publications, this research thesis defines a research methodology for detect sapm email based on user behaviour with the help of NLP, and delivers a proposed new email security framework using NLP pattern classification to reduce the risk of business email compromise (BEC). The title of "Detecting business email compromise and classifying for countermeasures.", is chosen to represent the core of the research and the deliverable. The research question, "What patterns are detected in email attacks on business systems?", is selected to guide the research project and to focus on the potential AI NLP contribution to email security.

The thesis is structured into the following sections. The formalities and chapter 1 provide a concise introduction to the research and the findings. The main body of the thesis starts in chapter 2 where the relevant literature is systematically reviewed to extract up-to-date information on email, email attacks, and the most recent AI research for protecting emails. It also gives gaps for research and the

possibility for researchable questions and testable hypotheses. Chapter 3 continues the literature review but focuses concisely on four published research reports from scholars who have done similar research. This information instructs methods and processes others have used to do similar research. From this knowledge I set up the methodology and testbed for this project. Chapter 4 reports the results that hold the evidence to answer the research question. Chapter 5 tests the hypothesis and answers the research question based on the results of chapter 4. There is also an attempt to reconcile the findings with the literature gaps identified in chapter 2. In chapter 6 the thesis is concluded by acknowledging the research limitations and the openings for further similar research projects. A full list of references follows, and the Appendix contains the python code built for creating the NPL model to analyse the spam emails and to calculate the confusion matrix metrics.

Chapter 2

Literature Review

2.0 INTRODUCTION

The literature review started by searching related past research works and published articles in various academic journals in the AUT electronic library. Furthermore, the preliminary search started by searching generic topics that related to business email compromising (BEC) such as “Business email compromising”, “business email attack and protections”. Then after that the search expanded into detailed topics related to BEC and artificial intelligent email protection plus using semantic engines for email protection. The searches were systematic and the leads from one search helped to guide the next in both the AUT electronic library, and the Google scholar search engines. In the preliminary search it was found there were a limited number of research articles in BEC from the social engineering perspective. However, after finding articles they were indexed, and items were saved to google drive for further analysis and review.

Today, email is the main business communication channel. It can be widely used common domains (google, Microsoft, amazon) or private domains. However, these email communications are threatened by various vulnerabilities, such as hackers, business competitors or from government spy organisations. BEC failure can be devastating for any organisation. Failure and compromise can cause damage to an organisation’s reputation and impact company profits and threaten in a worst-case example bankruptcy for the company. Attackers use email as an entry point to the corporate networks, and to compromise email users through social engineering attacks. Social engineering techniques are the most common techniques used by attackers to earn trust from email users to open or read the email, and then to infiltrate the corporate network. Once users open malicious emails it can spread to the whole network or else it can open back doors to hackers who can then lurk and monitor all the corporate network. Spear phishing is another method cybercriminals are using via the business emails. Spear phishing emails target individuals or user groups within specific organisations to spread malicious malware or mislead the users to compromise information (Spear phishing, n.d.). Section 2.3 reviews these BEC risks and potential vulnerabilities.

Most corporate users who use business email systems have insufficient knowledge about cybersecurity, email security or phishing attacks. Also, most email security breaches are common because of users' lack of education about basic cyber security knowledge or else corporate users' negligent actions (Kessel & Allan, 2015). In addition, attackers always trying various techniques and tools to compromise the business email, therefore security researchers have stated that the psychological states of the email users and age are crucial matters for BEC (Schenkman, 2020). Consequently, in section 2.3 of the Literature review these matters will be analysed. Additionally, the legal support and safeguards to protect from attacks are often inadequate. Cybercriminals are well organised, often based in several countries, and act internationally. Therefore, criminals can escape when or before being caught by authorities because of a lack of international laws and legal barriers to cooperation between countries.

Business organisations cannot depend on outdated security methods and tools because criminals are always trying different tools and techniques. Consequently, in this research, the artificial intelligence methods of natural language processing are investigated to evaluate the help they can give to prevent spam emails and harm. Natural language can process the content of an email and evaluate the intention of a sender. It considers the headings, content words, reference patterns and the links. Therefore, the techniques for the use of the semantic engine are found in the literature for identifying phishing email attacks (Verma & Hossain, 2014). Section 2.3.1 reviews the natural language processing (NLP) engines, its techniques, limitations, and detection methods. The following sections report the literature review of email architecture, email compromises, email attacks, BEC impacts, and natural language engines.

2.1 BUSINESS EMAIL INFORMATION SYSTEM ARCHITECTURE

Since email was introduced to the business world, it has become one of the most popular and most used applications on the Internet. With the help of continuous research and development, email communication systems have significant architectural changes and are now mandatory infrastructure services for the global communication system. Today most of the modern email systems are based on the "store and forward" message model. However, in the early stages of the email communication medium the sender and the recipient both needed to be online when the communication happened. Nonetheless with the new store and forward message

model once the sender sends to the recipient, an email will deliver to the recipient's email server and store there until the recipient opens the email. However, with cloud computing most of the email servers are replaced by the cloud email providing services.

Email architecture contains three major components, which are the Mail User Agent (MUA), the Message Transfer Agent (MTA), and the Message Access Agent (MAA) (Fahad , 2020). A user usually uses MUA to compose an email message and through a webmail interface or client email software such as Outlook. Then this message transfers to the MTA which receives messages and sends out messages from the MTA. The MAA is responsible for pulling the messages from servers to the client programme. In the following sub sections, each component of email and the architectures are reviewed in detail.

2.1.1 Anatomy of an email

The constructed components of an email message are important for BEC, because email users can clearly identify the legitimate emails, and ignore phishing emails by overlooking particular email components. Therefore, in this section the major components of any email message and responsibilities of each component are reviewed. Email contains three major components, those are the envelope, the header and the message body.

The first component of an email message is the envelope. The Email envelope is similar to the postal envelope, it contains the sender address and recipient address. However, the email envelope contains all the communication records for the client and the SMTP Server (Figure 2.1).

```
S: 220 foo.com Service Ready
C: HELO bar.com
S: 250 OK
C: MAIL FROM:<james.baker@xyzinc.com>
S: 250 OK
C: RCPT TO:<mary.jane@abcinc.com>
S: 250 OK
C: DATA
S: 354 Start mail input;
C: Actual email is sent here
C: .
S: 250 OK
C: QUIT
```

Figure 2.1 Email Envelope

Figure 2.1 shows an example of sample envelope C: representing the email client and S: represents the email server. To start, the client and the server handshake with the HELO command, and the client starts the communication with the MAIL FROM command that contains the users email address (Costales, 2002). In addition, if in the event the email is not delivered to the recipient this value will be used to send it back with the non-delivery report (NDR). After that the client sends the RCPT 10 command which tells the server the recipient email address. In addition, only one RCPT command can be used for one user, but it can use multiple RCPT commands for multiple recipients.

The email header is critical to correct email messaging. An email header contains mandatory information to deliver the email message which includes the sender's email address, recipient email address, the date the email is send out, and the email subject. The email header also carries time stamps of when the email was send out by the sender, and all of the MTA received times as well. Therefore, when any user wants to understand the MTA transfer date and times the user needs to read the email header from bottom to top (What is an Email Header?, 2020). The header also provides the email routine information as well. Normally when an email is sent from one point to another it will transfer to several MTAs, and each time it transfers to one MTA agent to another, then the MTA agent email will keep the record of the time stamp of when MTA received that email (Costales, 2002). Furthermore, to send out an email, the email system needs to have at least one client side MTA and vice versa to receive an email message, the server must have at least one MTA. These features by the MTA will help to trace and determine SPAM emails.

Email protocols are not a part of an email however they are an important matter for email communication. Therefore, the next section reports three common email protocols that are widely used.

2.1.2 Email communication protocols

A protocol is a set of rules that need to be followed for email communications. To maintain the communication standards and to protect the email security, email servers need to follow specific protocols that are defined by several organisations such as the Internet Engineering Task Force, IEEE, or ISO. In this section three major protocols that are used in email communications are elaborated. Those are Internet Messages Access Protocol (IMAP), Post Office protocol (POP3) and Simple

Mail Transfer Protocol (SMTP). The SMTP protocol manages all the email message communications between the servers, and POP3 and IMAP protocols manage the email message sending process from the server to the email client (Holtz, 2020).

2.1.2.1 SMTP

SMTP email protocol was introduced in 1982 and updated with new rules in the year 2008 by the IETF (Bazydło, Lasota, & Kozakiewicz, 2017). This protocol is commonly used to send email from one server to another server as mentioned above, and it cannot receive email messages. SMTP is an application-level protocol and text based as well. The SMTP protocol uses two types of methods to deliver the messages, which are End to End, and the Store and forward method. The End-to-end method is used to send the email messages to different organisations, for an example, Microsoft to Gmail. In contrast the store and forward method is used to communicate within the same organisations. When the SMTP server receives an email, it keeps a copy of all received email until the server gets the delivery notification from the other server (Vaibhav & Chandan, 2020). Unlike other protocols such as POP3 and IMAP, the SMTP protocol does not need any sort of authentication to process tasks. In every SMTP server it has two specific ports for communications, namely, Port 25 and port 465. Port 465 is encrypted by using SSL/TLS encryption and is SMTPS (Edgaras G, 2020).

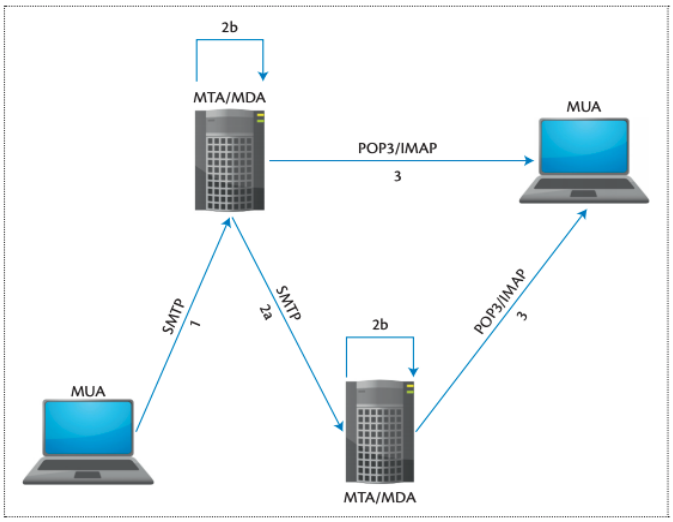


Figure 2.2 SMTP email communication

(Bazydło, Lasota, & Kozakiewicz, 2017)

The user provides the email message and the recipients of the email address to the MUA then the client MUA passes that email message to the dedicated server via a

SMTP protocol (Bazydło, Lasota, & Kozakiewicz, 2017). However, if the server is not available or if the MUA could not send the message to the server, it will send it to another relay server to send out the email.

As shown in Figure 2.1, the SMTP server greets the client with a 220 command and then the client sends an EHLO or HELO command to the server. When the server has acknowledged the mail client's greeting, the server sends a 250 command to say the server is ready to serve the client. The client then provides sender information via the MAIL FROM command (Bazydło, Lasota, & Kozakiewicz, 2017). Once the server has accepted the sender details that are provided by client, the server issues a 250 command to the client for the next step. As a final step in this process the client sends the message contents via the DATA command. This whole communication happens under the Internet Message Format standards (Bazydło, Lasota, & Kozakiewicz, 2017).

2.1.2.2 POP3

Unlike the SMTP protocol, the Post Office Protocol version 3 (POP3) needs authentication, which means this protocol manages authentication by using user credentials. However, the POP3 protocol is responsible for downloading email messages to the client machine and the protocol can only communicate one-way which means server to client. Once it downloads the email message from the server to the client computer, the POP3 protocol will delete the original message from the server. POP3 mainly uses two ports for communication which are port 110 for non-encryption communication and port 995 for encryption with SSL/TLS (commonly known as POP3S) (Edgaras G, 2020). POP3 protocol was famous in the dial up Internet connection era because it allowed to user to work in an offline mode and to do the synchronisation with the server when the user connected the client computer to the Internet. Eventually POP3 became unpopular because of several disadvantages such as, if a client device lost the user connection, then they may lose all the email messages as there are no copies of the original email message in the server.

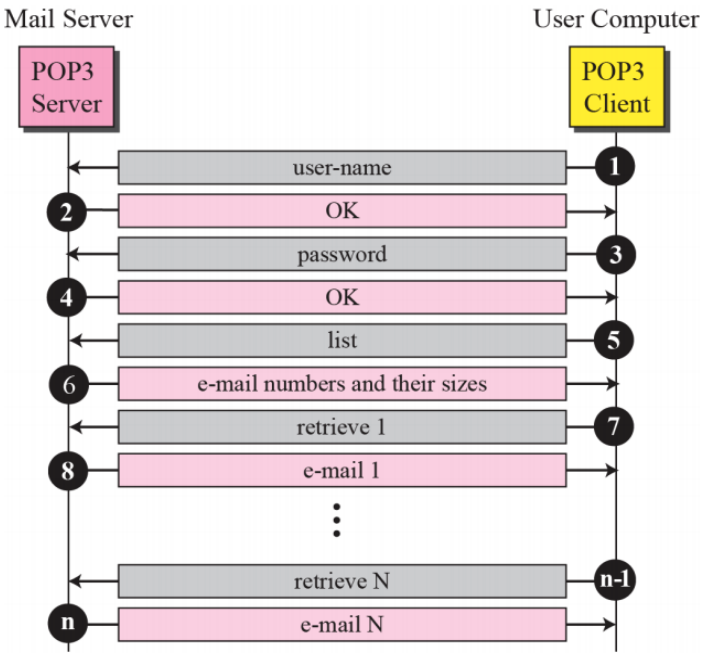


Figure 2.3 POP3 email communication (Fahad , 2020)

Figure 2.3 shows how the POP3 email communication process happens. First the POP3 client provides the username and password for an authentication process. Once the server has accepted the authorisation details the POP3 client sends the LIST command to get the number of emails and their sizes. Once the client gets the necessary information it sends a RETR command followed by an integer value to get the email messages from the server. In addition, once the client gets the nth email message it also sends the DELE command, followed by a particular integer value to mark it for deletion from the server. Once all the messages are received from the server, a POP3 email client sends the QUIT command to end the session with the server.

2.1.2.3 IMAP (Internet Message Access Protocol)

The first version of the IMAP protocol was introduced in 1986 and then it was followed by the second version, introduced in 1988 under the request for comments number 1064 (Kozierok, 2005). However, the third version of IMAP was introduced in IMAP3 because of expert dissatisfaction with the overall outcome from IMAP2. The RFC 1203, IMAP3 was published in 1991 with the title of “Counter Proposal” (Kozierok, 2005). Later IMAP2bis and IMAP 4 were introduced respectively. Compared to the POP email protocol, IMAP has several differences. For example,

the IMAP protocol allows a client email agent to handle email messages in the server without downloading them to the client's computer. Secondly, the IMAP server keeps the copy of the original message, and then multiple clients can access the email message from the server and there is now no need to worry if one device is lost or the original message is deleted in any client. As IMAP is a bi-directional communication channel, if the user reads an email message from one client it is marked as read in the server. Finally, the IMAP protocol uses the two port numbers which are the same as the POP3 protocol. Port 143 is used as a default port which is non-encrypted communication and port 993 is used for encrypted secure communication.

2.2 BUSINESS EMAIL COMPROMISE IN THE MODERN CYBER SECURITY DOMAIN

The current business world is decentralised, mobile, and more competitive compared to a few years ago. Instantaneous communications always connect with everyone, investors, governments, industry experts, media, clients, and suppliers. Business organisations need to exchange data and information quickly as the modern business world is highly volatile and based on the information exchange strategy. Therefore, from small coffee shops to multibillion dollar valued companies have adopted Internet communication and the cyber world to exchange their information. The digital economy concept drives modern businesses pushing them into the cyber world (The digital economy, new business models and key features, 2014). The adoption has numerous advantages such as mobility of business function, worldwide operations, acquired new technology, reaching more clients, and competitive suppliers. With the increased advantages information security threats present disadvantages and negative risk to businesses. The next two paragraphs review business email compromise risks in modern businesses.

At present, the Internet is the most powerful and important tool for communications, therefore lots of people and businesses connecting with each other via the Internet and they exchange an incalculable amount of business-related data and information. The data and information is valuable and worth large amounts of money plus business secrets disclosure can seriously damage the business. As a result of the technology revolution and inexpensive computing power immoral

people get to steal business information from the Internet or from the business information system illegitimately. In 2015 the US FBI has issued (Internet Crime Complaint Centre) three public announcements that related to BEC and those businesses compromised caused losses of us$179 million (Zweighaft, 2017). The incentive for large returns drives hackers to steal company information more and more. Nyakanyanga (2020) states that in South Africa business organisations have lost more than R2.2 billion each year because of cyber-attacks and Internet frauds. Furthermore, on the report she has stated one specific company has lost nearly 40 terabytes of company data resulting in the company's share price dropping 4.7% because of the cyber-attacks. The threat is a not unique to one single country and for that reason business organisations need to take extra effort to increase their data protection awareness and levels. Figure 2.4 reviews these risks.

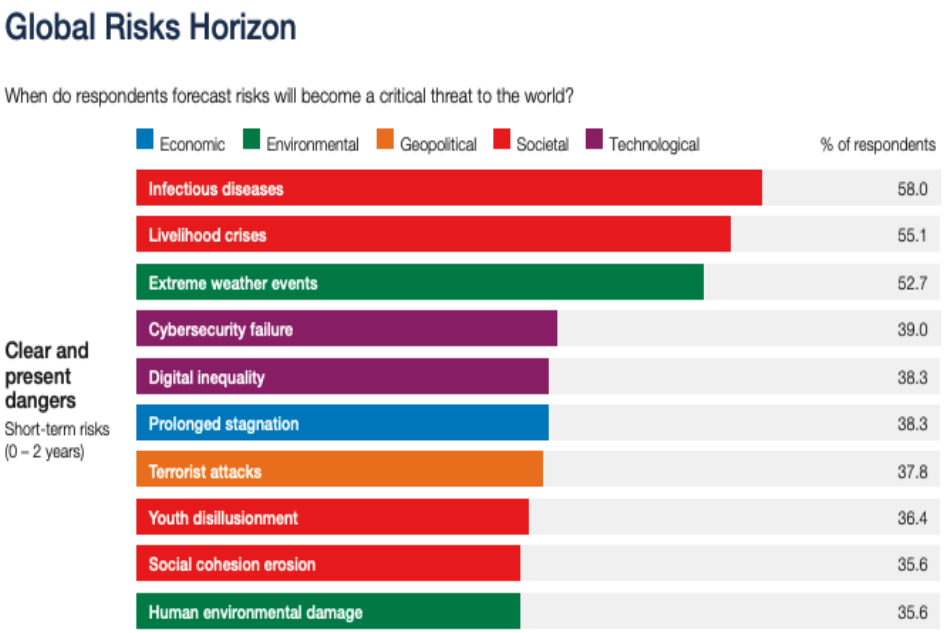


Figure 2.4 (The Global Risks Report 2021 16th Edition, 2021)

The Global Risk Report 2021 published by the World Economic Forum has cyber risks on top of the ranking. Furthermore, with the recent COVID-19 pandemic most of the workers and business engagements occurred online. Consequently, the sudden adoption to online work and often from insecure home networks, there are security issues for the businesses. The speed at which governments locked down, the countries and security experts did not get enough time to analyse potential threats to their business and organisation communications networks. Rosenthal and Oberly (2020) stated that cyber criminals adjust their strategies according to the COVID-19

pandemic settings. In the report it is stated that the US FBI has received approximately 240,000 complaints regarding BEC. Additionally, most of the users of cloud services assume the cloud security will protect against cyber criminals. However, the US FBI has issued a warning BEC still can happen on cloud-based email services or other cloud based services such as file sharing, calendar, instant messages, and so on (Rosenthal & Oberly, 2020). Consequently, all the above examples and facts suggest cyber security should be more proactive than reactive.

In the previous paragraph cyber security is the biggest threat to the business world security, hence governments and business organisations are investing heavily to protect data and information from cyber criminals. Hendry (2020) has stated that the Australian federal government is planning to spend $270 billion dollars in next 10 years period to increase the existing cyber security strength and implement new systems. At the same time Australia and New Zealand have launched a joint programme call CyberCX which is targeted to improve the NZ cyber security domain (Knowles, 2020). CyberCX NZ programme CEO has states that each year in New Zealand, business organisations become victims to cyber-attacks and loses of millions of dollars (Knowles, 2020). Investments in cyber security are to increase during the next few years as more and more cyber security threats are identified and predicted. However, all of these investments can be meaningless if organisations or the system users not aware of the BEC or the threats that can come via emails. This means all the staff or the members who are using the organisation network need to be educated regarding the BEC and also the threats that can come via fake emails plus how to avoid them. In the next section the importance of human factors is discussed for enhancing the cyber security discipline.

2.2.1 Human factors and BEC

Unlike machines, mistakes can be expected and happen regularly with any human. However, in cyber security human mistakes should be minimise as much as possible (as it is impossible to prevent human errors). According to Greenberg (2015) human mistakes and errors are the main reason for data breaches and security vulnerability issues. The 2015 report published by National Security Agency also stated human behaviour is a complicated issue in the cyber security spectrum because computers follow instructions, but human behaviour is always unpredictable (Science of Security (SoS) Initiative Annual Report, 2015). Most of the time cyber criminals are using social engineering for BEC and attack the organisation employee’s behaviour.

Pepper (2020) stated that BEC is getting extremely costly because of unpredictable human behaviour. In addition, the problem is growing because most security policies and protocols are developed to prevent actions are based on common human behaviour, however if in case of odd human behaviours or mistakes, then all the security enforcements implemented in an organisation can be breached.

The 2016 US presidential election was competitive and tightly fought. While both the parties are busy with the election campaign processes, hackers were implanting malware to the Democratic National Committee (DNC) office computers (Mak, 2017). As DNC is a major political party in US politics it has a high level of security physically and electronically. According to the ABC news website Russian hackers send 29 phishing emails that targeted the DNC campaign and all of them had impact (How Russia-linked hackers stole the Democrats' emails and destabilised Hillary Clinton's campaign, 2017). One staff member opened one phishing email that was sent by cyber criminals and information from DNC party's email servers was stolen. The staff member action compromised the whole security system in the DNC information system, and attackers stole 30,000 emails and wiped-out original emails from the email servers (Satter, Donn, & Day, 2017). This is an example that emphasises human factors and human behaviour as critical in for BEC.

When cyber criminals cannot get to the target information remotely via phishing email or malware, they try to compromise an inside employee to be a malware courier. Attackers try to emphasis an employer's displeasing side or offer bribes to deliver their malware to the organisation's information system. This kind of behaviour is hard to detect, and it could sabotage the whole business organisation, therefore Michael and Eloff (2020) have suggested business organisations should monitor employees email communication and social media activities to understand about their behaviour. Van and Allan (2015) state that attackers can be joined to business organisations as an employee to implant malware for further attacks or else steal information for selling to the competitors. An ordinary security system cannot detect human behaviour and alert for the further attacks, thus, to improve or avoid cyber security vulnerabilities Michael and Eloff (2020) suggest identifying the threat as early as possible and be alert on suspicious for employee electronic activities.

In Summery, human behaviour is a critical aspect in BEC that requires security planning and management. Therefore, in any cybersecurity framework it should be mandatory to consider human behaviour and psychology. Cyber security

defence education and new machine learning security systems can increase the security level of BEC and control of human behaviour. However, there are no assurances for cyber criminals also using artificial intelligent to harvest human behaviour and do advance social engineering to plan BEC. Consequently, the best way to prevent BEC is to continuously filter and monitor behaviours.

2.3 VARIOUS ATTACK TYPES USED FOR BUSINESS EMAIL COMPROMISE

In any information system Confidentiality, Integrity, and Availability (CIA) are the important elements because if one of these is not achieved by the information security system then the business organisation is vulnerable to failure. The CIA concept is also valid for any email system, therefore whoever oversees the email security system should closely check the CIA points to make sure they are covered in the implemented security system. In contrast most of the time email messages are transmitting in untrusted networks over the Internet once it leaves the main server. Hence, every email message needs the implementation of proper security enforcements to prevent email security compromises (Stine & Scholl, 2010). Since the email system is widely used for communications, the cyber criminals are targeting business email frequently and aggressively. Nevertheless, to stop BEC, security experts need to know the common threats for the business email system. This helps them to mitigate the email attacks and guides them to implement and manage effective security systems to protect the email system. Therefore, in the following sections I will analyse several attack types that cyber criminals are using to compromise the email security.

2.3.1 Phishing Attacks

Phishing attacks are the most common attack type cyber criminals are using against business email systems. A phishing attack is not only technology, but it also involves a percentage of social engineering for the targeted user or group of users. Attackers are using emails, instant messages, SMS, and other social media instant messages to gather legitimate users' information and send a phishing form. Phishing messages can trick knowledgeable users as well because they look almost like the legitimate messages. Phishing emails are also used to install malware into the targeted users' computers. According to Tandale and Pawar (2020) there are three phases of phishing attacks. The first phase is to deliver the phish message to the targeted victims and the second phase is to encourage victims to act. Mostly the action is to

click a link to direct to a malicious website or install malware or reply to that email with information. The last and third phase is selling the stolen data or demanding money from the victims (Tandale & Pawar, 2020). Tandale and Pawar (2020) has classified modern day phishing attacks into two main categories, which are those that exploit email-based phishing attacks, and web-based attacks. For exploit attacks attackers target vulnerabilities in web browsers and force users to install malware plugins that steal data. In the web-based attacks the user will be redirected to a similar kind of well-known fake web page, this can be similar in appearance, like a bank account web site or similar to other important web sites such as email, social security or tax claim web sites.

2.3.1.1 Spear Phishing

Phishing is a generic method or terminology that addresses the practice of distributing hundreds or thousands of emails by targeting victims to penetrate the information system to steal the data. In contrast, Spear phishing is distinctly targeted to one person or a specific user group in the organisation. Therefore, before the spear phishing attack, perpetrators will study the user(s) background thoroughly from social media, or Internet searching or from other electronic media (Gendre, 2018).

Figure 2.5 Difference between spear phishing and phishing email (Chivers, 2020)

As shown in figure 2.5, spear phishing emails clearly address a targeted user by name, unlike phishing email that mention the sender's name. What is more, the body of the email look the same as a legitimate email from the real sender, therefore it is a tedious task to identify the different between a spear phishing email and a legitimate

email. For that reason, O'driscoll (2020) has stated that spare phishing is the most dangerous phishing attack compared to other phishing attacks.

The Ubiquiti Network is a computer network service providing company which is based in the US. In 2015 this company was the victim of a spear phishing attack and the company lost $40 million (O'driscoll, 2020). In the beginning the attacker send an email to the Chief Accounting officer instructing to do a series of money transfers to acquire another company shares in secret (Vardi, 2016). Unfortunately, the company did not realise this was a fraud until the US FBI notify them and Ubiquiti Network stopped further money transfers to the criminals (O'driscoll, 2020). In another incident, hackers targeted the company Epsilon which is an online marketing company. Compared to the Ubiquiti Network attack this attack was a highly technical attack because the spear phishing email contained the link that downloaded malware to stop the antivirus software and enabled remote access to the host web servers (O'driscoll, 2020). Consequently, these two cases illustrated spear phishing emails dangers.

2.3.1.2 Other Types of Phishing Attacks

Whaling is like a spear phishing attack where criminals use almost the same technique to get victims. However, the only difference between whaling and spear phishing is that in whaling cyber criminals are targeting the powerful people in business organisations. The reason why attackers are targeting this level of individuals is because they have access to the organisation's sensitive data or else, they can do large money transactions without prior approvals. In 2016 a snapchat employee disclosed the company's payroll data to the perpetrators by replying to an email that looked like it came from the company CEO (Petters, 2020). Voice phishing or knowing as Vishing is another phishing attack which is performed via voice media such as call or VOIP. Most of the time vishing scams show they are representing a trusted company such as a financial institution, debt collecting or an outsource company who support Microsoft technical support (Owaida, 2021). SMS Phishing or else Smishing is a text message-based phishing attack method. Cyber criminals send SMS or text messages that contain a URL to users to redirect to forged web sites which look like legitimate web sites. Sometimes the links will download malware to the device and infect to the network, when mobile devices are connected to the corporate networks.

2.3.2 Spam

A Spam definition is the distribution of unsolicited email messages in bulk. Spam emails are used for marketing purposes however cyber criminals use this method to distribute malware or harmful contents to email users. In addition to the distribution of malware through Spam emails there is financial motivation for cyber criminals to distribute the Spam emails. According to Asif (2019) team spammers earn around 3.5 million US dollars every year (Karim, Azam, Shanmugam, Kannoorpatti, & Alazab, 2019). Approximately 236 billion emails are exchanging every day and 53.6 % are spam emails. Thus, it shows how much spam email are trafficking over the Internet and why it is important to prevent the spam email traffic. In another report Okunade (2017) has stated the US FBI has reported us$12.5 billion lost from business email users because of Spam email and their affects.

Similar to phishing emails Spam emails are used as a method of delivering malware or else invite the user to perform an action such as download software, install a plugin to a browser or click a link to redirect to a fake website. The other biggest risk is when a business organisation email is compromised to the spammers, spammers can use those compromised email credentials to generate more spam email. Therefore, government or other authorised organisations can question the Spam email generation from legitimate email domains. In a worst-case example Spam can shut down the Internet and connections from Internet service providers (Types of Email Attacks and The Damage They Can Cause, 2016).

2.3.3 Social engineering attacks

Social engineering is an art that criminals use to manipulate peoples' behaviour and actions. These manipulations can involve psychology and technology (Lord , 2020). Social engineering is chosen by the cyber criminals because it is difficult to detect human mistakes or human behaviour plus it is easy to gain the access credentials of users rather than cracking passwords for access from outside of the organisation (Kenton, 2019). Heinbach (2020) has listed four stages of social engineering attack. In the first phase an attacker will investigate the target and collect information, as much as possible, such as personal information, personal pressure points (such as debts, legal issues, and so on). Then in the second phase the perpetrator will start to gain the trust with the victims by using the gathered information via messaging. In the third phase the attacker will deploy their harmful attack and collect the necessary

information, this may be credentials, financial transaction details or business intellectual property. The perpetrator targets anything valuable for the business competitors or financially valuable objects. In the final stage the attacker will remove the forensic details that can be traced back to them and move to the next target (Heinbach , 2020).

In 2020, Barbara Cocroran, host of the famous Shark tank TV programme was scammed of nearly $380,000 by social engineering techniques. Principally in this instance the attacker studied her investment properties and her personal assistant email information plus how the money was handled in her organisation. The attacker had well-crafted the scam email like a legitimate one, and used an email address the same as her personal assistant email but one letter was misspelled (Sandler, 2020). Toyota Boshoku Cooperation is one of the largest auto part suppliers for Toyota cooperation. In 2019 that company was scammed nearly US$37 million dollars. The attacker influenced the financial authority person and had them do a wire transfer to the account information in the scam email (Mathews, 2019). These examples show social engineering is a danger as the electronic technical system could not prevent human behaviour in the information system. Therefore, the best way to protect is to prevent spam email coming to the legitimate user email accounts.

2.3.4 Malware Attacks

Malware or “malicious software” that uses any type of software or programme code delivers in practice malicious intentions (Malware, 2021). Cybercriminals use malware in different formats to compromise legitimate communications. In most instances they send malware as Trojan horses, virus, ransomware, or spyware. However, the common goal behind the malware is to get access to the systems or monitor the system illegitimately. Today most email users are alert and will not open executable files which come with an email as attachments, however, to deliver the malicious payload cybercriminals also use different methods such as fake business document or fake URLs that look like the genuine ones. These download plugins or documents create BEC (Email: Click with Caution, 2019). Common users trust antimalware or virus guards for protection. In contrast the truth is antimalware can only protect the system if the malware is identified prior to an attack. Without knowing the malware signature before attack the antimalware cannot protect the system (Broadhurst & Trivedi, Malware in spam email: Risks and trends in the Australian Spam Intelligence Database, 2020).

Ransomware uses a resilient encryption algorithm to encrypt a victim's computer data, and the attackers demand money to release the private key. Common algorithms used are the Advanced Encryption Standard (AES) 128 –bit or Rivest-Shanir-Adleman (RSA) 2048 –bit encryption (Broadhurst & Trivedi, Malware in spam email: Risks and trends in the Australian Spam Intelligence Database, 2020). Usually, a ransomware file contains the payment instructions. The attack commonly distributes on the Internet as an email attachment. As soon as the user downloads the attachment the ransomware encrypts the computer hard disk drive and demands the money payment. Secondly, Trojan horses are also commonly used malware to attack business emails or the corporate networks. According to Roderic and Harshit (2020) most commonly the Trojan horse file comes as a doc.js or pdf.js and this file contains the URL that can download the necessary executable files and DLLs to compromise the information system. In addition to these malwares and viruses, spywares and adware are also risks. The Checkpoint web site lists another 5 types of malwares that cyber criminals are using to compromised systems and BEC. Those are Crypto mining malware, Mobile malware, Botnet, Info stealers and APT Trojans (The 5 Most Common Types of Malware, 2020).

2.3.5 Directory Harvest Attacks

Directory harvesting attacks (DHA) have the main goal to gain information about all e-mail addresses used in a domain, and it can be a denial-of-service effect if the email volumes are high. Attackers use brute force attacks or else exhaustive key searches to harvest business email information. With the help of modern technology attackers generate different permutations to generate email addresses and send bulk email addresses to the domain. After successful entry the email is delivered to the matched legitimate email address. The attacker will then get the delivery notification and can be flood spam emails to the email server (Bencs´ath & Vajda, 2007). DHA may not directly harm the business email, but it can expose organisation domain information and the email formats that company is using. By exposing domain details and other network details, other security vulnerabilities in the business organisation are exposed.

2.4 VARIOUS ISSUES AND PROBLEMS BECAUSE OF BEC

When cyber criminals target more and more attacks on business emails, it causes many problems for the business organisations. Therefore, in this section various

issues and problems that arise because of BEC are identified and discussed. In addition, related incidences based on case scenarios are used to understand the importance of business email security, and the urgency to prevent the BEC.

Five major types of BEC methods that defined by the US FBI are reviewed. Those are CEO fraud, False Invoice scheme, Account compromise, Attorney Impersonation and Data theft (Zorz, 2019). In the CEO fraud scheme, the attacker acts like a company CEO and requests to make financial transactions from a person who has authorisation to do. This person can be CFO, Chief Accountant or may be a Chief Financial Controller. This fraud is also known as a Business Executive Scam, Financial Industry Wire Fraud, and Masquerading (Security 101: Business Email Compromise (BEC) Schemes, 2016). In the Account compromise scheme the attacker hacks a legitimate email account and sends multiple requests to vendors to make the pending payments to scammer's account. The False invoice scheme or Fake Invoice scheme is also known as the supplier swindle scam usually targeting overseas suppliers for wire transfer of funds to an attacker account. The Attorney impersonation scam is when the attacker shows as a lawyer or legal representative to request sensitive data or do a quick wire transfer to a fraudulent account (Security 101: Business Email Compromise (BEC) Schemes, 2016). In the Data theft scheme scammers do not request any funds but rather personal identification information about executive level employees. Hence, attackers will use the information for CEO fraud attacks.

Because of BEC business organisations are facing innumerable problems, and I shall categorise these into three sub sections in this review. Those are Economic, Legal, and technical problems.

2.4.1 Economical problems and Issues

Most of the business organisations are spending considerable amounts of money to protect their data because of increasing cybercriminal activities. In the case of any data breach regarding privacy law the penalties are high. It does not matter to hackers and cyber criminals how much security reinforcement or legal reinforcement is implemented against the cybercriminal activities, they continue trying to breach information systems to gain financial benefits, steal data, or information. When a system breach happens there are some direct costs involved such as fines from government authorities, security experts or consultants' fees, legal fees or paying

damages to clients as well. However, when BEC and data breaches occur there are hidden cost also involved such as share price drop, damage to company reputation and key employee morale. This hidden cost and indirect costs are hard to recover, and it can impact the organisation in the long term. Korolov (2019) states that according to the Association of Financial Processional (AFP) survey 81% of companies received fraud emails and 54% of them have lost money. 29% percent of the companies have confirmed they lost more than $100,000.

The invoice scam is another method that hackers are using to scam business organisations, and this is a type of BEC (Volkman , 2019). Swanston (2021) has stated on the ABC news web site that a former employee of the National Australian Bank stole au$4.2 million dollars with an invoicing scam. The incident shows even high security organisations such as banks are not secure with BEC, because of human behaviour. However, damage recovery processes and remediation for other hidden damages were possible in this case. Another incidence was reported in the Australian Cyber Security Centre annual report where an Australian consulting firm was tricked to send $240,000 to a fraudster in Malaysia. In this case the company CEO's email was compromised and criminals used that compromised email to trick the CEO when in Malaysia. These are strong case examples that show that cyber criminals use social engineering techniques, and monitoring for awareness of company CEO's moves and who is working in the financial department (ACSC Annual Cyber Threat Report July 2019 to June 2020, 2020).

Any business organisation needs to plan for damage recovery and have methods ready if any damage happens to the organisation. Mostly the potential damages are estimated in financial values and insured with insurance companies. However sometimes insurance companies will not cover BEC damages under the damage claims. Hence, companies may not have any recovery method for financial lost because of BEC. Tollefson (2020) has illustrated even some organisations who has a cybersecurity insurance policy it does not cover BEC attacks necessarily. The case example is the Virtu Financial BEC incidence. In that incidence the insurance company refused to pay the loss of $6.9 million as it was an employee action rather than authorised access of the information system (Tollefson, 2020). Consequently, these are examples that show BEC is a serious issue and can have inordinate financial losses, even in organisations that have already implemented precautions to recover the losses.

Financial loss is a direct consequence of BEC. Also, reputation loss is associated with BEC. Consequently, BEC has long and short-term impacts on a business organisation. Therefore, enterprise email security needs extra levels of protection in addition to ordinary cyber security system protections. Potential for security failure must be factored into business organisation planning to recover unpredictable financial and reputational losses.

2.4.2 Legal problems and Issues

Most BEC deliver significant data breaches and these data breaches cause immediate consequences and long-term consequences. Fines, forensic investigations and new security reinforcement implementation are a short-term consequence. On the other hand legal issues can be recognise as a long term consequence. The reason is that it takes time to recognise and to calculate the damage of BEC. Also, how far it impacts the business organisation stake holders, such as clients, public shareholders, and community. In the 2017 PWC report it is stated that consumers believe business organisations are proactive about their sensitive data protection security when they shared their data with the business (Consumer Intelligence Series: Protect.me, 2017). The General Data Protection Regulation and commonly known as GDPR act implemented in 2016, protects the data privacy of European Union citizens and residents (Palmer, 2019). Therefore, if a hacker gains access to any EU residents data and compromises the user privacy then the business organisation who process those data need to pay heavy fines, and be faced with other legal consequences such as prohibition of business or to continue the business under probations (Felman, 2019). BEC in these jurisdictions has a significant impact and a direct effect for the business operation.

Insurance is a powerful tool for risk mitigation in business processes. These risks can be data, financial or other natural disasters. However, when it comes to data risk, insurance is a complicated issue. That is because a data security breach can lead to different kinds of losses and those losses may not be covered in an agreed insurance policy between the victim’s business organisation and the insurance company. An example is the lawsuit between the Federal Insurance Co and Ameriforge Group Inc. The insurance policy holder which is Ameriforge Group Inc was scammed by a CEO fraud which is a BEC. However, the insurance issuer, which is Federal Insurance Co, denied the claim acceptance because the agreed policy was not covering CEO fraud or other methods that cyber criminals are using to scam

business organisations. In other cases, insurance companies have argued that they are not covering spoofing email scam damages. Consequently, BEC can lead to complicated legal issues if BEC is not specifically named, and proactive steps taken to prevent it.

2.4.3 Technical problems and Issues

After compromise of the email system, business organisations need to handle technical problems apart from legal and economic problems. The reason behind this is for a fully digital forensic analysis within the organisation network and other digital systems as well. Cyber criminals plant malwares that can enable backdoors to access the organisation's system again and forensic investigation is required. As a first step the technical team needs to gather accurate evidence and to stop leaking data still happening on the system. These processes must be done quickly and effectively, and not to cause or disclose disruption of networks or systems. It can impact an organisation's full capacity of business functionality and can cause other problems as well such as customer trust and legal issues.

When a ransomware attack occurs it encrypts all the data that is exposed to the ransomware, and the organisation might need to pay money to decrypt the files. In some cases, after data breaches organisations might need to replace the whole system. An example in Australia, cybersecurity specialists recommended to the Tasmanian Government to replace the ambulance communication system to prevent further data leakage after a BEC incidence happened (Cooper, 2021). The problem was that transmitted data was not encrypted because of it is hard to transmitted encrypted data in remote areas and critical information was compromised. The cost of the BEC data breach was a whole new system for millions of au$. Modern business organisations rely on other service providers for technical services such as infrastructure or content management services or communications systems. If these third-party service providers are attacked, it effects many business organisations who are getting their services from them. In that case the business organisations need to replace their connections and integrations to another service provider, or they might need to switch to another technology. These kinds of migrations are not simple to do because technical experts need to analyse requirements of integration with the new service providers, and they are expensive.

In summery BEC can lead to many problems as discussed in the above sections. They can be legal, economic, and technical issues. Nonetheless, all these issues and problems can harm the business in bad way. Therefore, business organisations need to address BEC seriously and they must increase the security of the organisation email system. Many of these issues can be mitigate by introducing stronger security system on business email system, and better education of the users to make them aware of current dangers.

2.5 NATURAL LANGUAGE ENGINES

Artificial intelligences are being implemented into email security systems to improve monitoring, response time to attacks, and big data management. BEC is driving an increase in security hardness, training, and extra layers of security protection on email communication systems. There is a move to beyond traditional security methods and to expert systems and Artificial intelligence (AI) to increase the system security capability (Bulao, 2021). Bulao (2021) has stated that by using AI technology in security systems it helps to improve thread identification and prediction. At the same time by using AI it helps to avoid human mistakes that could be lead to cyber security breaches. The changes are a huge achievement for the cyber security industry because as explained in section 2.2.1 human mistakes are the leading factor for BEC. Deep learning is a subset of AI and it has advanced algorithms which are similar to the human brain learning curve. Musthaler (2016) has stated that deep learning AI has helped to identify the malwares which are attached on emails and if any unknown files contain risks. Another advantage is that AI based security systems can detect fast and can response to the threat quickly, therefore it will minimise the threat to the targeted information system.

Natural Language Processing or commonly known as NLP or the NL engine is a subset of AI that is a trending technology to fill the gap of security weaknesses for social engineering attacks. Compared to other AI technologies NLP is heavily involving with protecting against BEC, because NLP uses different algorithms, logic, and different models to understand human language (Benishti, Can Machine Learning Help Prevent Business Email Compromise?, 2022). NLP can analyse the contents of the email and can decide whether it is spam or not based on the trained it is given. This process is like a well-trained technical person reading individual emails and marking them spam email or not. An example of NLP use is virtual

assistants such as Siri, Google assistance or Alexa from Amazon. Hence, next section it will review the concepts and anatomy of the NLP.

2.5.1 Anatomy of Natural Language Processing

In simple terms NLP means a computer system can understand the human languages and understand the communication intent. This communication channel can be word based or may be a voice channel. However, NLP can be categorised in to two parts and those are Natural Language Understanding (NLU) and Natural Language Generation (NLG) (Khurana, Koli, Khatter, & Singh, 2017). Figure 2.6 shows the broad classification of NLP according to Khurana, Koli and team (2017).

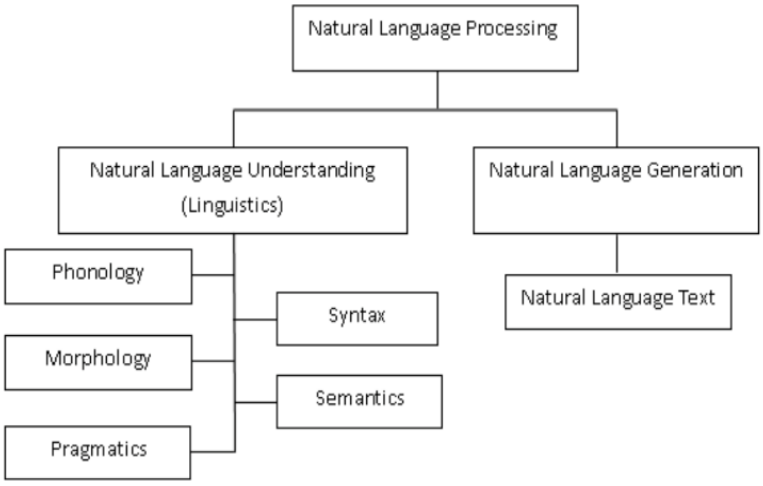


Figure 2.6 Classification of NLP

(Khurana, Koli, Khatter, & Singh, 2017)

Natural language understanding is a linguistics science and has five problems that need to be addressed according to Khurana and the research team (2017). The five components have their own tasks. In languages there are words that have the same sounds but different meaning therefore when it comes to NLP computer interpretation it needs to understand exactly what words are suited to the situation according to the input sound. However, for Email security systems it may not need a deeper level of understanding because almost every BEC is in plain text. The second problem is when dealing with NLP, is morphology, which means breaking a word in to prefix, root and suffix. Reshaping is a good example for understand morphology. For this word “re” is prefix, “shap” is root and “ing” is the suffix part. Thus, when

the NLP engine processes the words, it needs to understand distinct words in small parts (Khurana, Koli, Khatter, & Singh, 2017). Pragmatic is the problem of need to understand more words based on where they are used for the same meaning. To address this problem the NLP engine needs more word knowledge and their usage plus when those words are used. The word itself suggests syntax, refereeing, and arranging of characters that make the word make sense. This can be special characters or may be letters. Furthermore, this problem is directly related to when NLP is used for email security, because some spam emails do not use the same words as it can be detecting on security spam filters. Cyber criminals are using different character organisation like the word to by-pass security barriers. When humans hear words, they try to understand the words with the context, however this approach will not help when it comes to the computers. Consequently, NLP needs different approaches to handle this problem and it applies semantics analysis (Khurana, Koli, Khatter, & Singh, 2017).

This paragraph will discuss the different significant phases of NLP. There are six main phases commonly practising now in NLP. Those are Morphological analysis, lexical analysis, syntactic analysis, semantic analysis, disclosure integration, and pragmatic analysis (Figure 2.7).

Figure 2.7 Phases of NLP

(Banerjee, 2020)

Figure 2.7 shows the phases of NLP action. First the NLP engine tries to understand the characteristic of words as stated in the phrase morphemes. Thus, the engine will try to understand the word by its most minor parts as it is trained. The biggest challenge in the morphological analyser is that each language has its own way of dividing the word into prefix, root, and suffix. However, in the morphological model,

there are several implementation characteristics available. Those are the dictionary lookup model, fine-state morphology, unification-based model, and functional model (Bikel & Zitouni, 2012). Lexical analysis is the second phase of the NLP engine. In this phase, the NLP engine analyses the structure of words and separates the portion of text into words, sentences, and paragraphs. According to Banerjee (2020) lexical analysis has two normalisation practices which are stemming and lemmatisation. Stemming means removing word suffix by using a redundancy-based algorithm, and in contrast, lemmatisation is the process that makes the procedure of locating the root of each term (Banerjee, 2020). The third phase of the NLP analyses the sentence grammatical structure and builds the relationship between the words logically. This phase is commonly known as syntactic analysis and is sometimes confused with semantic analysis, which is the next phase of NLP phases. Banerjee (2020) has stated in the semantic analysis phase that the NLP engine has to analyse the linguistic interaction between the words and defined the real meaning of the sentence. The discourse integration phase explores the meaning and structure of the sentence and tries to build the connection between the words. The last phase, which is pragmatic analysis, tries to match and analyse what it tries to communicate through the real world sentence. This phase might consider the time, location, and other external factors based on the message context.

2.6 CONCLUSION

The literature review in this chapter has provided comprehensive knowledge about BEC and the different attack models for business emails. Section 2.1 has focused on components of email and necessary protocols related to business email communications. The modern cyber world is rapidly growing, and it is alarming the impact of cyber-criminal activities over the Internet. Therefore, business organisations need to be proactive to protect their email communication channels from cyber-attacks. Research has discussed various factors that can lead to BEC, and the human factor is the primary factor among them. A computer system works according to the given instructions, but for humans, it depends on various factors such as emotion, knowledge, culture, and many more facts based on the situation. Therefore, the human factor is the most vulnerable point in BEC as discussed in section 2.2.1.

Section 2.3 has focused on various attack types that cyber criminals use to compromise the business email system. Mostly they used human factor focused

attacks such as phishing, spams, and social engineering attack types. Commonly attackers combined two or more attack types to compromised email systems. However, malware, phishing attacks, and social engineering attacks are the most famous, and those attacks are still breach email security barriers because of human errors.

Because of BEC business organisation can face unlimited problems, these problems can be long-term and short-term. Conversely, this complication can be categorised into three major categories: legal, economic, technical issues, and problems. Legal issues and problems can differ based on where the business operations are happening; however, economic issues and technical issues may affect the business organisation directly when BEC happens. With the BEC, it can harm a company reputation and lead to financial issues plus it can destroy the business organisation. At the same time, BEC may cause several big issues such as having to replace the current system, or complex technical upgrades and ransomware data solutions. Security researchers have focused on AI-based security systems to prevent BEC and to resolve all the problems mentioned earlier. NLP based security approaches are getting popular, especially for avoiding spam and phishing attacks to email systems. Section 2.5 has discussed the NLP and its anatomy, including the standard 6 phases required for NLP.

In the chapter 3 the research problem and methodologies to be used in this research are defined. It also discusses similar studies which are relevant to this research and the set up for the experimental work.

Chapter 3

Methodology

3.0 INTRODUCTION

Chapter 2 critically reviewed a wide range of literature regarding the BEC. In addition, the reviewed literature provides a guide to learn about essential components of email and standard communication protocols that email uses to deliver messages. Chapter 2 also covered various factors that can cause BEC and different attack types that cybercriminals use against email systems. It is also mandatory to understand why BEC is essential for business organizations in the modern cyber world. Therefore, the literature review also examines three categories of significant problems that business organizations may deal with because of BEC. The main purpose of chapter 3 is to develop research requirements that distinguish the challenges presented in chapter 2 into researchable issues and problems for laboratory testing.

Reviewing and analysing previous studies positions new research, and helps design the research hypothesis, research question and set up the research environment. Consequently, section 3.1 is allocated to deeply analysing five similar studies about BEC and AI-based counter security systems. These five similar studies aim to advance the domain knowledge about setting up an accurate and efficient test environment, selecting correct tools, and collecting accurate data to answer the research question and sub-questions. The research hypothesis and research question are defined in section 3.2. Section 3.3 focuses on data gathering methodologies and sanitizing the collected data for analysis. Section 3.4 outlines the limitations faced by the selected constraints, and the obstacles and challenges expected in doing technical testing. Overall chapter 3 gives the proposed methodology for this research project.

3.1 REVIEW OF SIMILAR STUDIES

To understand and build the research methodologies for this project, five completed similar studies are analysed to learn from the methods, questions, and challenges others have found. These selected previous research publications will help to understand what kind of research methodologies need to be implemented to archive this research project objectives and to identify the obstacles before they arise while researching in the laboratory environment. The selected research publications cover a

wide range of technologies and research methods about the BEC and proposed countermeasures that use AI methodologies to prevent the BEC attacks.

3.1.1 E-Mail Spam Classification via Machine Learning and Natural Language Processing

Junnarkar, et al., (2021) has studied 5 machine learning algorithms to analyse the email contains and out of five they have shown the Naïve Bayesian algorithm and the Support Vector Machine based algorithm is more accurate for spam filtering processes. In that research paper, the researchers have reviewed literature about URL analysis, NLP, and ML approaches. Furthermore, Junnarkar et al. (2021) has evaluated the Random Forest machine learning algorithm used for text classification. In any algorithm, it is a vital factor how quickly it can respond and move through the learning curve (Junnarkar, Adhikari, Fagania, Chimurkar, & Karia, 2021). Furthermore, it has been stated in the research paper literature review that Neural Network-based algorithms are not suitable for spam rejection and showed less performance on NN training. In contrast, the Naïve Base algorithm has shown more performance for the isolated spam filtering plus low processing resources to reject the spam emails (Junnarkar, Adhikari, Fagania, Chimurkar, & Karia, 2021). Finally, researchers have discussed different spam filtering formulas developed by Outlook, Gmail, and yahoo companies to prevent spam traffic (Junnarkar, Adhikari, Fagania, Chimurkar, & Karia, 2021).

As stated in the previous paragraph, Akash et al. (2021) have explored five core machine learning algorithms in their research paper. Those are:

- K-Nearest Neighbour
- Naïve Bayes
- Decision Tree
- Random Forest
- Support Vector machine

In the first phase of the research, the authors have done text classification known as text categorization or text tagging. This process is relatively common in natural language processing. The primary purpose of the text classification process is automatically classified into predefined categories according to the different requirements. Feature engineering, Gensim packaging, Pre-Processing the data, and Vectorizing the data are the four steps that the authors have defined in their text

classification process in this research (Junnarkar, Adhikari, Fagania, Chimurkar, & Karia, 2021). Generally, Statistics, Optimization, and mathematics are the three core principles based on every ML algorithm (Junnarkar, Adhikari, Fagania, Chimurkar, & Karia, 2021). Therefore, ML algorithms are not self-sufficient to learn themselves to understand unstructured data. Hence, the research team needs to structure the data for processing. To sort unstructured raw data, the researchers have employed a feature engineering method for creating vectors that cans easily understand and follow the ML algorithm. Generate Similar or else known as Gensim package are used for topic modelling. After categorising and context analysing Akash et al. (2021) have sanitised the text. This has a sanitising process named as Pre-Processing the data which removes all the HTML tags, accented characters, special characters and stop words. In the Vectorizing procedure researchers have word2vec models and by using this model the team has achieved prediction of a similar meaning sentence or similar words (Junnarkar, Adhikari, Fagania, Chimurkar, & Karia, 2021).

To ensure confidentiality, integrity and availability of the email system, the research team has focused on filtering the URLs contained in the emails. In this URL filtering process, they have sub categorized four steps as listed:

- URL un-Shortening
- URL blacklisting
- Presence of trigger words
- Identification of special characters

(Junnarkar, Adhikari, Fagania, Chimurkar, & Karia, 2021)

Hackers are always trying to trick legitimate users by hiding the malicious URL under a fake URL looking like legitimate URL. However, to detect these URLs the research team has used a URL un-shortening method to discover the actual URL and procced for further processing. In the second step of the URL filtering process, the research team has compared the un-shorten URL with blacklisted URLs and if any record is matched, the email is marked as a spam email (Junnarkar, Adhikari, Fagania, Chimurkar, & Karia, 2021). Thirdly, Akash et al. (2021) have focused on detecting if the URL contains any trigger words that are contained in blacklist data sets. If any words match a URL in the blacklist data set, then the system marks that email as spam or malicious email. Sometimes malicious email URLs contain some words that are detected in previous spam emails. Therefore, if the comparison engine finds any word or word phase that match to existing spam words, it lists that email as

spam email by the spam filter engine. Every language has characters which look like symbols or a combination of characters that show another character. For example, in the English language it contains I and | or w can show as combined with two v s like "vv". Hackers use this tricky point to mislead legitimate email users to a false website to steal information or download malicious payloads. To avoid this loophole, research team has implemented an extra step to detect any suspicious special character identification process in URLs that an email contains. If the scan engine found any malicious objects in the four steps outlined, then the scan engine will mark that email message as a spam email. In a final stage, the research team used OR logic to combine the result in text classification and the URL filtering process results. Consequently, if any result is positive to spam classification, then that email message is marked as spam.

3.1.2 A Systematic Review on Spam Filtering Techniques based on Natural Language Processing Framework

The modern cybersecurity world is competitive and agile because creative hackers are always trying to find an ambiguity in the security system to compromise the information system behind them. Therefore, researchers are always trying to implement and introduce new security measures to the information security market. As discussed in chapter 2 spam is the biggest annoyance for legitimate email users. It might be the easiest way to distribute malware to any information system as it needs human interaction. The best way to stop spam email is to detect them before arrival to the user inbox, therefore, it needs a mechanism to identify spam emails from genuine emails coming from genuine senders. Consequently, Garg & Girdhar (2021) has focused on spam email filtration contrivances based on NLP plus filtering different spam techniques and messages. Those are Appending, spam that uses images, blank spam, and blackscatter spam messages (Garg & Girdhar, 2021).

Around the world, there are more than 6000 languages worldwide; therefore, filtering spam on each language is near impossible because each language has its own unique patterns, combinations, and meaning based on order of the words (Garg & Girdhar, 2021). In this particular research paper, researchers have reviewed spam filtering techniques such as: problematic recognition method researched by Shami, AdaBoost; a method that by Carerras which is more effective that the Bayesian filter method; and, a Machine learning algorithm based spam detection method implemented by Mishne et al (Garg & Girdhar, 2021). Secondly, Garg & Girdhar

(2021) also reviewed image filtering techniques, including OCR-based text detection algorithm proposed by Yassen and the research team (Garg & Girdhar, 2021). Garg & Girdhar (2021) have an OCR based text recognition algorithm combined with a ML algorithm to increase the accuracy of spam filtering techniques. The focus on detection is to determine the email address, subject and contents substances, and finally the action of a spam email is the most unsafe activity in any spam email (Garg & Girdhar, 2021). Email is not the only way that hackers can distribute the spam contents. It can be spread via SMS messages as well. Therefore, Garg & Girdhar (2021) also considered implementing spam detection techniques to identify spam in SMS.

Apart from the techniques mentioned above, spam identification techniques and methodologies also include:

- Whitelist and blacklist
- Checking the mail headers
- Support vector machine (SVM)
- Bayesian analysis

Figure 3.1 shows Pranjul Garg and Nancy Girdhar (2021) proposed framework that they proposed and research against for filtering SPAM messages and emails.

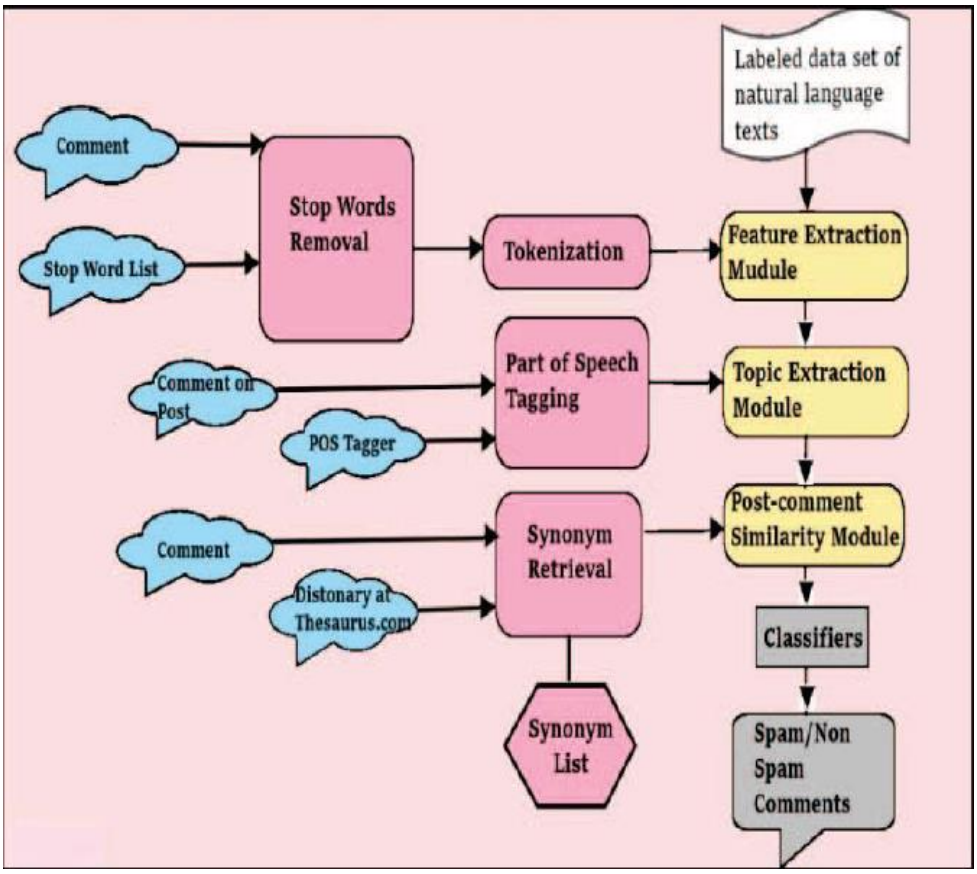


Figure 3.1 Pranjul Garg and Nancy Girdhar (2021) proposed a framework for spam detection.

Essentially NLP is a combination of NLU (Natural Language Understanding) and NLG (Natural Language Generation). NLU is the mediator between human language and the computer, plus from the other end, NLG converts raw data into human-readable format.

According to Figure 3.1, the researchers have treated the words by removing the stop words and comments. This process will make the process of NLU faster and more efficient. Then the engine tokenizes the words treated from the stop words removal process. Essentially this process will split sentences into a collection of words, and this is known as tokens. After tokenization, researchers arranged the words in to root structures, which can manicure, manipulate, and manufacture the meaning of a sentence or synonym (Garg & Girdhar, 2021). Lemmatization is a process that identifies and tags the words with the same meaning in different forms. By lemmatization process research team has increased the process speed of spam identification. Also, the POS tagger or else known as Parts of speech tagging also helps. Every language has some words where the way they are used can change the

meaning. The process of identifying these words is known as POS tagging and this process helps the NLP to identify the root form of the word form (Garg & Girdhar, 2021). The last step is the process that research team has defined to identify the signature of the spam. In this the process engine will compare all the histories of extracted words and the duration of the how long those words contained in emails messages stayed in the inbox. At the beginning of the process or when new emails are received it creates a list of new top words and according to the research team this list is one of the most important parts in the spam filtering engine (Garg & Girdhar, 2021).

Pranjul Garg and Nancy Girdhar (2021) used four steps to detect spam comments. Those are Profanity check, Pre-processing module, Topic extraction module and the sentiment analysis module (Garg & Girdhar, 2021). In the first step of the spam comment detection process the research team has checked the bad words with the previous generated lists and if there are any words found in history list those comments contained emails are marking as a spam. Then after that it comes to the tokenization process which is the tagging of the words and POS tagging as described in the previous paragraph. This process will take time as the engine needs to go through the entire email. In the final stage the proposed system extracts the topics found in the comments section. To archive this task Pranjul Garg and Nancy Girdhar (2021) have used the Latent Dirichlet Allocation model. In this process the sentiment analysis process can find the topics that related to comments (Garg & Girdhar, 2021).

For data analysing and evaluation metrics Pranjul Garg and Nancy Girdhar (2021) have used common evaluating formula that most other researchers used to report performance. The accuracy measure computes the ration of correct classifications to the total number of classifications (1). (TP= true positive correct malicious classification; TN= true negative correct benign classification; FP= false positive = incorrect malicious classification; FN= incorrect benign classification)

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

The precision of the algorithm is computed as the ratio of correct malicious classifications and the sum of correct malicious classifications and incorrect malicious classifications (2).

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

The recall or true positive rate is computed as the ratio of correct malicious

classifications and the sum of correct malicious classifications and incorrect benign classifications (3).

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

The F1 score is the harmonic mean for precision and recall and is computed in equation (4). The value of the F1 score is to acknowledge extreme values in a data set and hence moderate excessive theoretical normalization (P= precision; R= recall).

$$F1 = 2\ x \frac{P\ x\ R}{P+R} \quad (4)$$

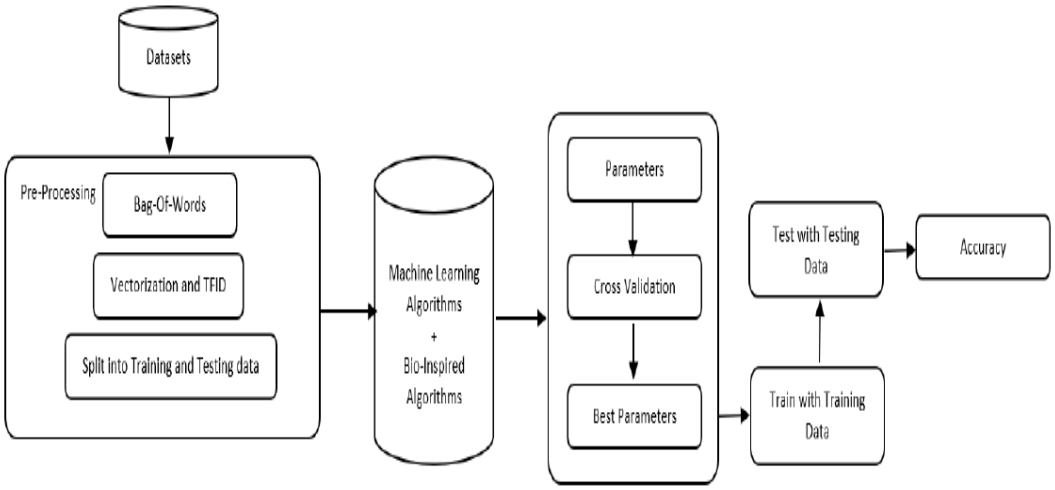
These formulae help to understand the how the spam filter engines perform and the accuracy. According to Pranjul Garg and Nancy Girdhar (2021) research results showed detection of Spam emails is high, however, harmful content is not easily detected. Pranjul Garg and Nancy Girdhar (2021) have stated a SVM algorithm showed 98% of accuracy in their research project. However, this performance is questionable as 2% failure is too high for the security expectations . A 2% chance of failure is more than enough to breach an entire information system if one spam email contains a malware payload.

3.1.3 Detecting Spam Email with Machine Learning Optimized with Bio-Inspired Metaheuristic Algorithms

This research paper has tried to identify legitimate emails from unsolicited emails. Combining Machine learning with Bio-Inspired algorithms, has the research team minimising the detection time and increasing the accurateness of the detection process for unsolicited emails. Gibson et al (2020) have selected Particle Swarm Optimization and Genetic Algorithms to categorise emails. By using these two algorithms the research team monitored the user email use and model improvements with parameter tuning (Gibson, Issac, Zhang, & Jacob, 2020). The research team has five objectives for this research:

- Explore machine learning algorithms for the spam detection problem.
- Investigate about algorithms in the acquired dataset.
- Implement a Bio-inspired algorithm.
- Test and compare the accuracy of base model with implemented bio-inspired implementation.
- Implement the framework using Python.

Figure 3.2 shows the proposed system overview by Gibson et al (2020).



(Gibson, Issac, Zhang, & Jacob, 2020)

Figure 3.2 The system proposed by Gibson et al. (2020)

The tools and techniques that have been used in this research project are as follows. This review will help me to understand what kind of tools can be used for this research project. Gibson and rest of the team has used the WEKA tool to understand the performance of the chosen algorithms and their functionalities for selected spam data sets. WEKA is an open-source tool that was developed by Waikato University which contains a collection of machine learning algorithms to resolve real word data mining problems. Therefore, this tool can be helpful for identifying which algorithm is most accurate for detecting spam emails and to optimise the parameters on each algorithm. Gibson et al (2020) have selected WEKA as a black box to choose the best algorithm for a spam detection engine. Furthermore, in their research model Gibson et al (2020) have used a supervised learning method which can classify spam emails into two categories after the training data and for the testing data. According to the research team, these two data categorisations improved efficiency of the model and it shows what parts or methods need more training (Gibson, Issac, Zhang, & Jacob, 2020).

They have used the Scikit-Learn (SKLearn) tool and KERAS API in their research project and shown how it will help in this research project. First,cikit-Learn is developed by using python and it was designed for statistical data analysis for those who do not have domain knowledge about statical data analysis (Pedregosa, Varoquaux, Gramfort, Michel, & Thirion, 2011). Gibson and the rest of the team has used SKLearn to train the “Fit” method, feature validation, feature selection and

parameter tuning of each algorithm (Gibson, Issac, Zhang, & Jacob, 2020) . For this research, the proposed method SKLearn tool will be helpful to understand how each algorithm (selected ones for proposed method) behave on task-oriented interfaces with supervision or without any supervision. Another most significant advantage is that python develops SKLearand therefore it is easy to integrate with any other tools that support the python ecosystem (Pedregosa, Varoquaux, Gramfort, Michel, & Thirion, 2011). Secondly, Gibson et al (2020) have used the KERAS API written in python. It also supports neural networks and consumes less CPU power, which is important because when AI tools consume lots of CPU power, the rest of the modules can get slow. Nevertheless, the proposed method will not use KERAS API as Gibson and the rest of the team used it in their research project.

In this project model training and testing phase is one of the most crucial parts. The research team has used known data and unknown data to train their models in the proposed system. Moreover, to get a high accuracy result Gibson et al (2020) have used the K-fold cross method. The research team note that there are weaknesses in the K-Ford method, which is test data can contain only spam emails which can mislead the training of the model. In contrast, Gibson et al (2020) have implemented a method to ignore this by cross-validation of the K-Ford method feeding data by providing a different set of spam and clean email data. As mentioned earlier the research team has used a bio-inspired algorithm to increase the accuracy of the machine learning process.

Another most important section in this research paper is the Bio-Inspired optimization algorithm optimization. In this process, Gibson and the research team have run through two approaches: particle swarm optimization and Genetic algorithm optimization. The Particle swarm optimization algorithm searches the best space for each particle in the data set. To archive, this Pyswarms library offers various calculations to optimize models and subset of data used for the training (Gibson, Issac, Zhang, & Jacob, 2020). By doing this optimization, spam algorithm can easily find the easiest root to find the pattern of the spam email within a short period of time. Therefore, as mentioned in the literature review, the spam filter should filter the email quickly and efficiently within a short period. In other words it is to find and tune the parameters for the elected Machine learning algorithm or a neural network (Gibson, Issac, Zhang, & Jacob, 2020). Following figure 3.1 shows pseudo code for how the research team has implemented parameter tuning by using particle swarm optimization.

Gibson et al (2020) have concluded that the Multinomial Naïve Bayes algorithm has shown high accuracy compared to other algorithms in spam filtering processes against the data set that they have used in their research project. In addition to train and test, Gibson and team have used several datasets including Ling-spam, SpamAssassin and Enron (Gibson, Issac, Zhang, & Jacob, 2020) . Apart from an alphabetic based data set, the research team has used numerical data set to test the spam email engine, PU1,2,3 and PU4 respectively. Moreover, in this research, the researchers have illustrated that a genetic algorithm can be more efficiently work with the PSO algorithm. Therefore the proposed method will use an algorithm parameter optimizer to boost the spam filter algorithm efficiency.

Figure 3.1 Pseudocode for particle swarming

```
Initialise Input Variables;
N ← No. of Documents;
X ← Datapoints;
y ← Target Inputs;
X_i = StopwordRemoval;
X_j = Vectorizer;
X_k = Tf-IDF;
Dict ← X_k1
Dict = Pre-Processed data;
Initialise ML Parameters; // This will include
    the key and the values
Declare ML Algorithm; // MNB, SGD, DT,
    RF and MLP
Def PSO:
    Initialise PSO parameters;
    C1 = 0.5; // Cognitive Parameter
    C2 = 0.7; // Social Parameter
    W = 0.9; // Weight
    N_i = NumberOfIteration;
    N_p = NumberOfParticles;
    Calculating the Dimension;
    (Key,Value) ← Parameters; // The
        parameters of the algorithm i.e
        MNB, SGD
    Call PSO G_Best algorithm; // Global Best
    PSO Module ← Dimension, C1, C2, w, N_i, N_p;
    Call Objective Function O_f;
    PSO ← O_f;
    Calculate the Best_Position of the Swarm;
    Best_Position ← N_i;
    Calculate the Measures;
    Measures ← Best_position, TrData, TeData;
    return Accuracy
Def O_f:
    for i < N_p do
        Initialise StratifiedKF;
        Calculate the Score;
        return The array of accuracies A_q
            // conducted with the
            dimensions and the Key and
            Value provided
    for t in test size do
        X_test and y_test = testing size;
        X_train and y_train = training size;
        Call the function PSO (training and testing data);
```

(Gibson, Issac, Zhang, & Jacob, 2020)

3.1.4 Classifying Unsolicited Bulk Email (UBE) using Python Machine Learning Techniques

This research paper has focused on a spam filtering method by using the spam-ham dictionary with the combination of different data mining algorithms by using a python engine. According to the research team, traditional spam filters can be classified into three different types: Word list, Blacklist and Whitelist, and finally Hash table (Mohammed, Mohammed, Fiaidhi, Fong, & Kim, 2013). As the name suggestions, the Word list contains special words that spammers commonly use to

attract the email recipients to click the malicious hyperlinks. If any email contains words from those word lists, the spam filter engine will mark those email messages as spam emails. The second method (Blacklist and Whitelist) contains a list of IP addresses that servers are generating spam email messages. Therefore, a spam detection engine will block all the emails traffic generated by Blacklisted IP addresses. Thirdly, some systems use hash tables. In this method spam filtering systems summarize incoming emails to a pseudo-unique value and compare them against the bulk emails. In this method, some spam emails can also get into a hash table. The spam filter may recognize the spam emails as legitimate emails. Therefore, to avoid outdated spam filters, business organisations urge accurate and fast spam email filters.

Mohammed et al (2013) have chosen python for their research project because as a programming language, python has a recognition that it shows good performance when it comes to data processing, text mining and string manipulation. In contrast, the research team also recognized that python has lots of text processing libraries and visualising libraries. Such as Genism, Plex, Scrapy, Dictionary and Pattern (Mohammed, Mohammed, Fiaidhi, Fong, & Kim, 2013). Mohammed et al (2013) states in their research project why they choose python as a tool because it contains more matured machine learning libraries compared to other programming tools. MILK, PyBrain, Elefant, and Orange are the few matured machine learning libraries available (Mohammed, Mohammed, Fiaidhi, Fong, & Kim, 2013).

In section two the research team has expressed how they build the Dictionary based spam classifier. In general, the dictionary is a basic data structure that every programmer uses to keep and contain with the keys. Therefore, it will help them sort or compare the contents or to find the specific element within the dictionary. Furthermore, the research team has mentioned that many spam word filtering dictionaries are available that can integrate with the python programming language, such as the E-Commerce Spam Triggers List compiled by Dr. Ralph F. Wilson (Mohammed, Mohammed, Fiaidhi, Fong, & Kim, 2013). In this research paper, the research team implemented a word list dictionary using a one key pair value. Also, they have identified this method is not the best solution to building a spam filter engine because of time to update the lexicon dictionary may take a week (Mohammed, Mohammed, Fiaidhi, Fong, & Kim, 2013). To overcome this obstacle the research team updated the word dictionary frequently and without waiting for automated updates.

Calibrating and Stemming are the two processes that help produce flexibility and adaptiveness in the spam filtering process within the proposed method by Mohammed et al (2013). In section 3 the research team is dedicated to describing their calibration process in the spam dictionary. In stemming its pre-processing, the email to transform all the words to their branches, the spam filter engine may miss synonyms and other forms of words (Mohammed, Mohammed, Fiaidhi, Fong, & Kim, 2013). The research team has stated that the stemming process is a straightforward process within the python engine. For python, there are many libraries available for stemming such as PyStemmer (Mohammed, Mohammed, Fiaidhi, Fong, & Kim, 2013). The stemming process is a good process that can apply to the proposed method in this research proposal. The reason is that the stemming process will help identify the stem words of different spam words plus this process can help improve the efficiency of the spam filtering process .

Conversely, spam email senders are using different techniques to bypass spam filtering. Sometimes they use special characters or dots or other techniques to represent specific words. For an example, Mohammed et al (2013) have mentioned in their research paper, spammers use different alternative ways to represent Viagra such as $VIAGRA$, V.I.A.G.R.A.S or ViAgrav to mislead the spam filter engines. To remove these special characters from camouflage words and produce the same word the research team has used regular expressions. While pulling the regular expressions together, the research team has tokenized the terms by using RegRx coaching module for further use.

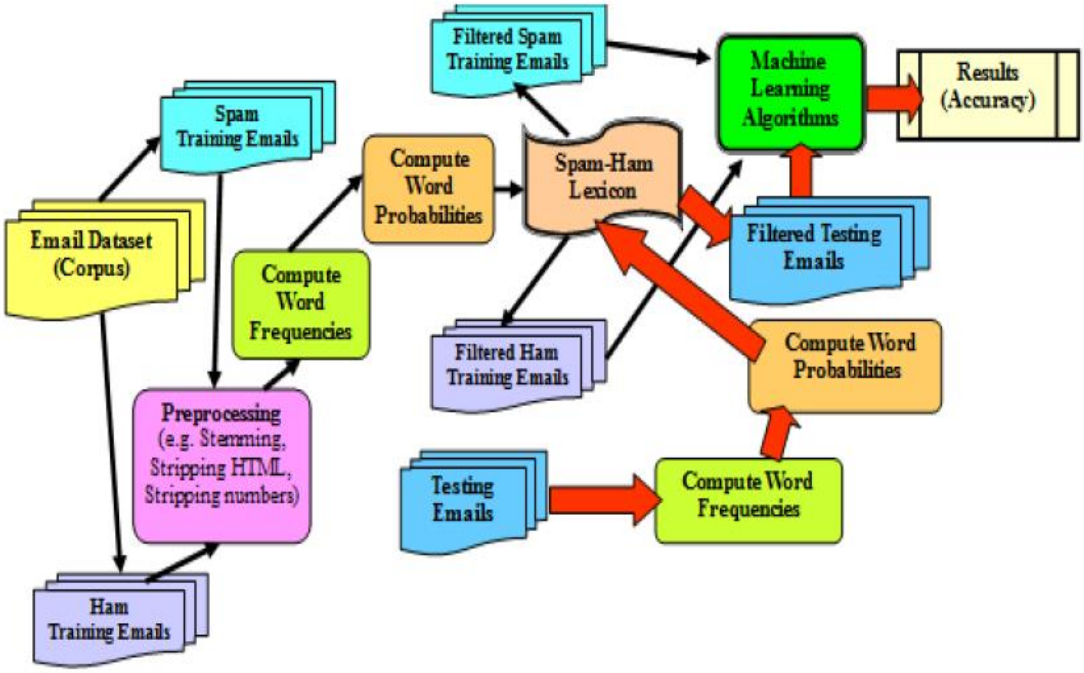


Figure 3.3 Mohammed et al (2013) proposed machine learning approach for the email classification.

The research team has produced a little lexicon form from each training that they provided to the engine; therefore they selected 32 words that can be found in spam emails and these highest probability words also can be used against the HAM filtering process as well. The research team has used a CSV file to import the word lists into the training process.

Mohammed et al (2013) explained the machine learning technique that they used for their research in section five. They selected an orange package written in C++. According to the research team this module is handy and quite easy to use for new machine learning algorithm training processes plus it contains a large variety of routines for data manipulations (Mohammed, Mohammed, Fiaidhi, Fong, & Kim, 2013). In the orange module are important functionalities necessary for machine learning processes, and it is easier to modify those functionalities by using python, including Attribute subset, bagging and boosting and decision tree (Mohammed, Mohammed, Fiaidhi, Fong, & Kim, 2013). The Orange framework also supports monitoring different classifiers for the data classification procedures such as a Support Vector machine, Naïve-Bayes and K-nearest neighbour (Mohammed, Mohammed, Fiaidhi, Fong, & Kim, 2013). Mohammed et al (2013) have figured out which Naïve-Bayes and Support Vector machine algorithms are best for spam and

ham email filtering by using this procedure. The argument is essential for this proposed research method because it can use the Orange framework to train the Naïve-Bays classifier for the proposed email filter method.

Finally, this research project accepts that the python engine and Orange framework are the best for machine learning processes for SPAM email filtering. Therefore, the proposed research will help understand what framework and what methods can be used for text mining. Furthermore, this research considers how to use different data sets to train the Nielson Bays algorithm and support vector machine algorithm. According to the authors the Nielson Email – 1431 dataset showed high performance with Nielson Bays and SVM against SPAM email detection. Therefore, this proposed research will use Nielson Email – 1431 dataset to train the spam email filter algorithm.

3.2 RESEARCH DESIGN

The previous section (section 3.1) has analysed four similar studies that provide guides to help formulate methods for this research project. Each researcher has shown how to design a research methodology and build NLP related tests for questions. In addition, reviewing previous studies also helps to fill the various researcher knowledge gaps and integrate their research methods into this research methodology. In this research thesis, Sub questions are essentially the same as the main question because sub-question (s) answers help to back up the main question. Therefore, in section 3.1, analysis also helps to explain sub question answers and to explore the answer more wisely.

In the next section (section 3.2.1) will summarise the review of the analysis done in section 3.1 that is beneficial to build the data requirements for this research project and filter out necessary points from the analysed research papers. Secondly, section 3.2.2 explains the research problem focus, hypothesis and research question, including the sub questions. The last part of the subsection (section 3.2.*) will explain the research phases of this research project and illustrate the data mapping. Data collection and data processing is a vital part in this research thesis, therefore in section 3.3 the data requirements and data processing steps using NLP engine are defined.

3.2.1 Summary of Similar Studies

Akash et al. (2021) classified spam emails using five major ML algorithms. Over those five major algorithms, Akash and the team have proved that Naïve Base theory-based algorithm plus Support Vector Machine based algorithms are more accurate for the spam filtering process. Apart from emails containing text analysis the research team has expanded their research area to filter spam emails based on URL, Images, and special character combination words such as “vv” instead of W. In the modern day, cyber criminals are weaponizing with the latest technology, therefore cybersecurity researchers are always trying to introduce a new concept or else introduce solutions by combining the existing technologies. Garg and Girdhar (2021) have published a research publication for spam detection by focusing on the email address, subject, and attachments. Moreover, they extended their research into OCR and SMS spam filtering. In addition, they have separated stop words and comments before the email contents process in NLP. Therefore, NLU can process the word phrases faster and efficiently.

Secondly, Gibson et al (2020) conducted research to detect spam emails by using a machine learning optimized Bio-inspired metaheuristic algorithm. In this research, Gibson and the team have focused on reducing the detection time and increasing the accuracy of filtering spam emails. To identify the best algorithm for spam filtering the research team used the WEKA tool which is developed by Waikato university to resolve real-life data mining issues. Apart from WEKA Gibson and the team have used SKLearn and KERAS API for data visualisation and statistical data analysis. The research report has confirmed that the Naïve Base theory-based algorithm is faster than compared to other ML algorithms.

Finally, Bulk email is one of the main tools that is used in marketing campaigns, therefore filtering a legitimate email from a bulk email campaign is another challenge when it comes to email security. Because spammers are using a bulk email campaign to spread spam emails or unsolicited emails to corporate users. By addressing this issue Mohammed and team (2013) have conducted research to classify unsolicited bulk email using python machine learning techniques. Initially in this research Mohammed et al (2013) implemented a word list dictionary using one key pair value. However, after realizing this process is not accurate, they manually updated the word dictionary rather auto update. Finally, Mohammed and team (2013)

stated Naïve-Bayes and Support Vector Machine algorithms are the best two algorithms to use in the spam filter process.

3.2.2 Research question and Hypotheses

Chapter 2 has discussed various research papers that illustrate business email security threats. At the beginning of chapter 2, it discussed the general architecture of the business email which included what the email envelope contains, and various protocols that are involved in email communication. Moreover, chapter 2 also illustrates various causes for business email compromise plus various attack types used by cybercriminals for business email compromise. These various types of cyber-attacks help to understand building a countermeasures framework. Similar research works always help to identify the methodologies, technologies, and tools to build the research. Therefore, section 3.1 discussed previous research works that are related to this research. Some research topics covered several NLP algorithms and some only focused on one algorithm. Ultimately, the main research question is developed in the next paragraph with the sub-questions that support building the answer for the main question.

Research Question: What patterns are detected in email attacks on business systems?

This is the main research question for this research and the main goal is to identify the pattern(s) that hackers execute against business email systems. To find the answer to the main question the following sub-questions are added to the research questions.

(SQ 1) What type of attacks and methods do hackers use for compromising business email systems?

In here I will identify the various attack types by using categorial analysis. To accomplish this task, I need to build my own corpora for this research.

(SQ 2) How do hackers use social engineering techniques to influence business email users to open unsafe emails?

The literature review has reviewed several kinds of literature about how cybercriminals are using social engineering techniques and methods to identify the target. Therefore, to build the protective framework it is important to understand how they collect legitimate user behaviour data on the Internet and how they are using

those data to compromise business emails. Hence, by finding answers to this question I expect to understand the social engineering effect of the business email compromising.

Hypothesis 1 (H1)

The new Corpus will add at least 2000 category records but might still miss new spam or phishing email attacks.

3.2.3 Research Phases

This section will elaborate on the research phases that progress to find the answers for the main and sub-questions in this proposed research. There are five phases (except the literature review) proposed in this research as shown in Figure 3.4.

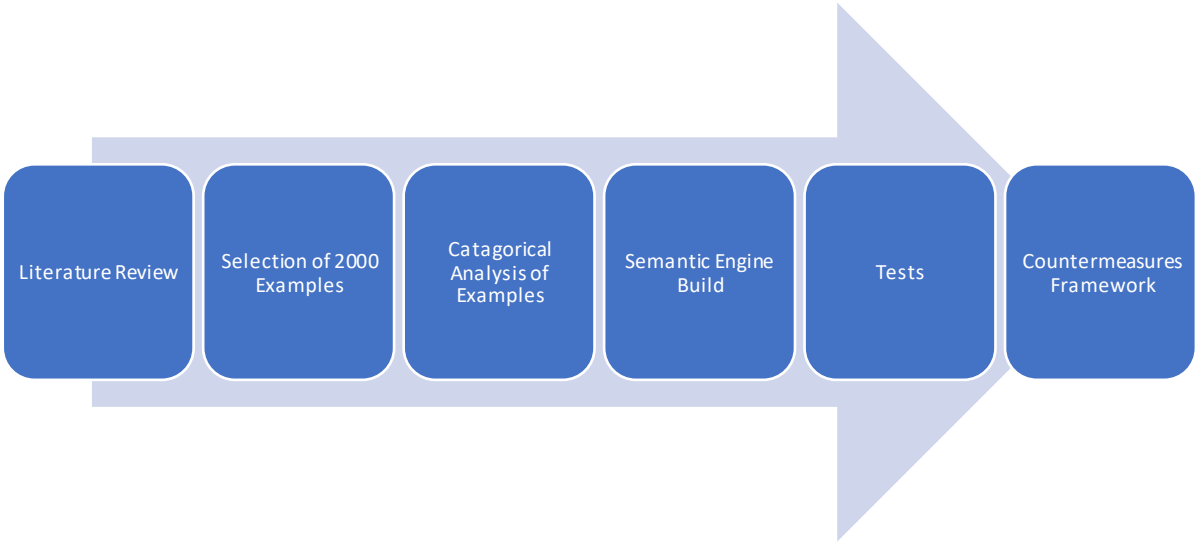


Figure 3.4 Research Phases for the proposed research

The first phase is selecting the examples of custom-made corpora for this research project. Corpora are a vital part for model training for identifying the affected emails that carry the payload to compromise the business emails. Nevertheless, for the new Corpora, it will use existing Ham and Spam corpora text. However, new corpora text will be added by using 2000 sample emails for phishing, directory harvesting attack emails and spear phishing email text categories. These texts will help to train the NLP model to identify illegitimate emails. To collect these emails, I will build a mock environment that has four computers and one mobile phone. Plus, it will be allocated freshly created emails for each device and by using these devices, it will search different types of targets such as medicine, finance-related products like

cryptocurrency, dating apps and online shopping. By doing this process I will expect to get spam emails to each device based on what they searched.

Secondly, the categorial analysis phase will process the categorical analysis by using the corpus created in the previous phase. This phase it is aiming to balance the data in each category. Otherwise, it can lead to false predictions and can cause the building of an erroneous model.

Before Applying the trained model to the email text, it needs to be sanitised and text processed. Because in English a single word can be given different meanings in different uses. For example, Bat, the bat can be a cricket bat or the animal bat. As computers cannot understand the logical meaning of words, we need to implement them. Therefore, implementing the semantic engine framework can extract the logical meaning of legitimate emails and can be used to identify actionable emails. After the semantic framework sanitises the email text it will test with the trained model and give a label and points. These points and labels will be stored for future analysis, and they will also use in the countermeasure framework to prevent email attacks.

In addition, the following network diagram will illustrate the plan for monitoring the network traffic to collect the search data to understand the social engineering attacks. Collecting the mock users' search details and network traffic details can map how many spams emails each user will get. By doing this process I can understand how attackers manipulate legitimate users to set up the business email attack.

3.3 DATA REQUIREMENTS

There are two types of data that need to be collected for this research. In the initial stage after setup of the simulated business network environment manually performs Internet searches. These searches will help to get more spam email traffic to the mock email address that is set up in the simulated device. The collected emails will be used to create corpora to train the model in the next phase. As mentioned in the previous section it needs to be a balanced collection of sample emails to train the model. Moreover, case-based scenarios will be used in this research project as cybercriminals can use different approaches and techniques to influence legitimate business email users.

After setting up the simulated environment, a dedicated Internet line has been provided for this simulated environment to reduce the external factors that can be affected by the process of collecting spam emails for corpora. This spam collection email process takes months because the need to collect spam emails based on what a user searches on the Internet. This collection will indicate how spam email senders are simulating the spam emails according to what a user searches on the Internet. In the next section, I will elaborate on what area-specific device searches on the Internet are used to generate spam emails.

3.3.1 Data Generation

Within the simulated network environment, Internet searches will be performed based on the following main categories.

- Sport
- Cryptocurrency and financial investment
- Online purchasing
- Adult entertainment (Games, Chat, Gambling)
- Medicine

Each device is assigned one category and will record all the inbound email traffic. To do the Internet search it will use a google search engine as a standard and all the search topics will be recorded.

3.3.2 Data Collection

Normally there are three types of data involving any system, those are inbound, stored, and outbound data. Though, in this research project, I will mainly focus on inbound emails. To emphasise how much it will affect the social engineering attacks for the email I will download search histories from the device browsers. As mentioned earlier there will be more than 2000 emails collected as inbound data and put in comma separated value format (CSV). In this CSV file, the first column will be renamed as the type that the categorised identified to train the model, such as Ham, spam, Phishing and social engineering. Plus, the other column will name as a text that contains the email text body to train the model.

3.3.3 Data processing

Section 3.3.2 described how to collect the data for this research and how they are categorised. In addition, in this section, it briefly described how to process the collected data for the analysing process. After building the corpora it needs to be tokenize hence, all the words from emails are collected. In the tokenisation process, it removes the stop words stemming and will remove the punctuations as well. In the Tokenization process, individual words will get a unique number to identify the sequence of the word with synonyms. Apart from that, the social engineering attack network traffic details will also be processed. The process will complete after collecting the user search data as shown in the next section. I will closely analyse this data to identify the patterns of spam emails and the relationship between user activity on the Internet.

3.3.4 Data analysis and presentation

This section overviews the analysis and elaborates on how the processed data is going to be presented. To understand spam email attacks it needs to be understood how many spam emails come to the user's inbox based on user activities on the Internet. Therefore, after categorizing the spam, ham, and phishing emails I will present the data in a heat map and also discuss the flow of each category. This finding will be used to implement the proposed countermeasures framework. Furthermore, I will conduct a comparative analysis based on the attack email category, for example, spam emails can get an unclaimed prize, therefore, how hard is it to understand for a legitimate user (Table 3.2).

Table 3.2 Attack level analysis 1

Category	Attack Load	Heading Title	Number of attempts	Difficulty	Search keywords
Spam	Link	Unclaimed prize	7	Moderate	Xmass, present
Phishing	Attachment	Gift card	2	Easy	Petrol, JB HIFI

As shown in Table 3.2 I will analyse the findings of how the attackers use social engineering to compromise the business emails. The aim of this analysis is to

influence the countermeasure framework to prevent the social engineering attacks on business email systems.

3.4 LIMITATIONS

This research project is focused on detecting the harmful emails that can compromise business email systems and to propose an intelligent framework to prevent them. When it involves machine learning (ML) there are lots of limitations that can distract the project. Therefore, this section explains a few general limitations that I faced because of data and scenarios.

Social engineering is a challenging issue to monitor and to prevent. Since, posting details on social media is unrestricted and it is always the human factor. This means the person to person interpretations can be different what they post on social media and what it means. In that case, when I collect the spam email, I only considered the users' Internet searches, not other social media details or posts on social media. Therefore, the end results do not completely reflect how social engineering attacks can compromise business email systems. In addition to the given time frame for this project, it is a tedious task to cover all the social engineering attack scenarios.

Secondly, in ML the success rate depends on how much we trained the model. To train the model it is mandatory to have accurate data to train the ML model. For this research project, I will build a corpus to train the NLP model, but there is no clear boundary or definition of how much data that is needed to train the NLP model. Therefore, the result can cause problems for some categories such as spare phishing emails. Also, some spam emails are targeting individual email users and this kind of attack does not reflect the entire picture of business email compromise.

Finally, this research project completely passed over the human factors that are involved with the business email compromise. Consequently, it completely assumed the end user will click the spam email or download the attack payload that is coming with the spam email. To automate this kind of process is a tedious task with a short time frame in test environment. As discussed in the previous paragraph an unmatured NLP model and the ignoring human factors need further research to be done to cover the full scenario of business email compromise.

3.5 CONCLUSION

Chapter 3 provides a comprehensive framework for this proposed research project. It has covered research methodologies, and research questions including the hypothesis and the data collection method as well. Similar research works are discussed in section 3.1. Four research reports are analysed of similar research projects that are relevant to this research proposal. Furthermore, using NLP to detect the spam email is a flourishing topic now in the email security research. Section 3.2 has the research main question, sub questions, hypothesises and research phases.

Also how it collect the spam email for an initial corpora to train the NLP model for this research project is described. Building accurate corpora is a crucial factor in this research project, therefore I am going to use more than 2000 sample emails to train the NLP model. To generate the social engineering impact, I have conducted various searches in the mock environment that is built similar to a corporate network. As this cannot be automated, I will be do these searches manually and record the incoming spam emails according to the users Internet behaviour.

Finally, section 3.4 has discussed the main limitations imposed by the setup of this research project. By identifying the limitations, it will give a clear idea about the value of the research findings, further research possibilities, and prevent misguiding the readers. In addition, this research might help future researchers to cover the scenarios better and fix the limitations discussed in the limitation section. In chapter 4 the research findings will be reported after using the defined methodologies and choices in this chapter 3.

Chapter 4

Research Findings

4.1 INTRODUCTION

Chapter 3 has defined the research question and related hypotheses, plus the sub-questions that support the main research question. Furthermore, section 3.1 reviewed several similar studies that helped build the research methodology and identify the technology and tools needed in this research project. Moreover, section 3.3 illustrated the data generation process, data collection methods, data analysing and processing to understand the business email compromise issue.

The main objective of this chapter 4 demonstrates the research methodology and present the research findings for business email compromise patterns as detected by the NLP AI application. Chapter 4 includes four major sections. Section 4.1 shows the variations involved in the data generation and data processing stages. The second part discusses (section 4.2) the research project test environment setup for collecting the spam emails to build the initial corpus. The next subsection (section 4.3) will explain the test cases plus the test case findings, followed by subsection 4.4 which is a summary to conclude chapter 4.

4.2 ENCOUNTERED VARIATIONS IN THE RESEARCH PHASE

Initially, it was planned to use 2000 sample emails to train the NLP model, however, after conducting several pilot experiments I have realised that 2000 sample emails are not enough to train the NLP model. The reason is there are lots of different types of HAM, spam, and phishing emails that are used to compromise business email systems. Therefore, it requires at least 1000 emails in each category to train the NLP model.

4.2.1 Data Collection

The data collection process is an important step in this research project because based on the first stage collected data will be used to train the NLP model and it impacts the success of the entire project. Before initiating the project, I planned to collect spam emails from my normal email accounts' junk folders, however, after setting up the test environment and starting to collect some emails on the test environment I realized the initial test data collection will not reflect the actual social

engineering effect. Therefore, the NLP model will not get proper training data. To avoid this problem, I have collected the model training data after setting up the test environment.

4.3 TEST ENVIRONMENT FOR RESEARCH

This research focused on business email compromising; thus, I have set up a small office network where devices connect wirelessly. I have neutralized this research network from other security enhancements such as intelligent network packet checking gateways, checkpoints, or security gateways. The reason is that this research only focuses on business email compromising and finding a solution based on NLP (Figure 4.1).

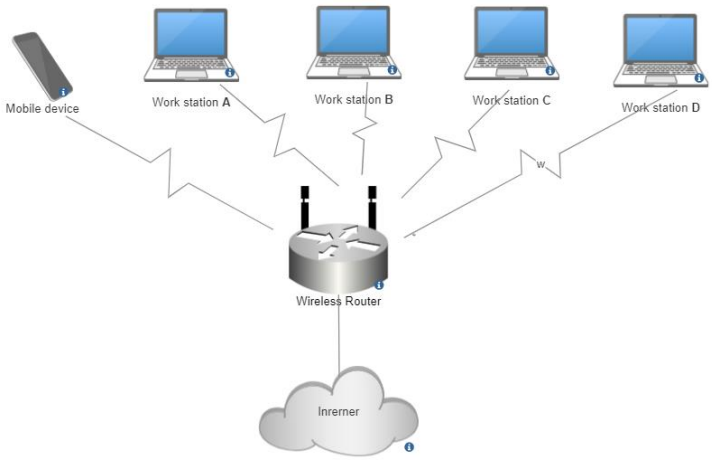


Figure 4.1 Test environment network diagram

4.3.1 Generating test data

As mentioned earlier, for training the NLP model it needs to build corpora, and to build the corpora it needs thousands of sample spam emails. To collect those emails, I performed several searches on the Internet which is shown in the following Table 4.1.

Table 4.1 Number of spam emails collected for model tests.

Related search areas	**Searched period**	**Total number of Spam emails in email Inboxes**

Investment opportunities-related topics	2021-01- 10 to 2021-03-31	372
Adult entertainment (casino, dating, traveling)	2021- 01-10 to 2021-04-27	874
Electronic consuming items	2021-01- 10 to 2021-04-27	412
Financial institutes and their products	2021-02-03 to 2021-04 – 18	279
Health products and medicines	2021-01-10 to 2021- 04- 10	422

In total, I have managed to collect just more than 2300 spam emails. However, after collecting these emails, I have realized most of the phishing emails contain pictures, not text. As this research proposal's primary focus is to analyse text-based spam emails and not image-based spam emails. Because it requires optical character recognition (OCR) and needs to filter the noises when you fetch the text from OCR (Yaseen, Abbas, & Ahmed, 2020). Therefore, in this research when it builds the corpora, I have selected Ham and spam emails for the main categories. The following image shows the constructed corpora.

	A	B	C
		Category	**Email Body**
		Ham	Take part in our marketing survey and Get $90 promo reward
		Spam	Take action TODAY - Up to 50% off on Premium subscriptions. Give it a go! It's time to fall in love. You deserve it,
		Spam	Play the Pets game to meet new people!
		Spam	This is to bring to your notice that you've been dealing with scammers over the years with regards to your long standing fund payment. You are advised to quit further comı
		Spam	Get your SmartMatch of today.Thousand of members get their perfect match on this site!
		Ham	Post Office Notification - Automated Delivery date: 9 November 2021 Your delivery status: On hold The courier assigned to deliver your package has dropped it off at your
		Spam	I'm attracted to guys with long hair. Leather jacket is a bonus. send a pic
		Spam	More People To Show Interest In CRYPTO In 2022
		Spam	€ 3000 welcome Bonus + 175 Free Spin Prices and products in effect November 12-17, 2021, unless otherwise stated. Prices are stated as of the date of this email. Prices a
		Ham	you have (1) package waiting for delivery. Use your code to track it and receive it. Schedule your delivery and subscribe to our push notifications to avoid this from happeni
		Ham	Well done Pubudu! Your email address was selected! Please make sure you claim your chance to receive this $2,000 PAK'nSAVE voucher. If you don't, we have to give this
		Spam	Fedex#44259, We've got a new message for you. We have sent you a message with the required information. Have trouble reading this email? Click here to open this email
		Spam	WhatsApp Missed voice message. Details Date: Dec 23 3:40 Duration: 09 seconds Play
		Ham	Dear Pubudu Gayan Delivery status: Cancelled Last delivery attempt: 45 minutes ago Current package location: Logistics Hub Courier Report: The address information was i
		Ham	Hi PUBUDGYAN, We are sure you appreciate our protection against dangerous viruses, hackers and cyber criminals. Unfortunately our protection has expired and we have
		Spam	African "Twist" Technique..Adds-6-Inches
		Spam	The Amazing Keto Supplement! Melt Fat Fast without diet or exercise. Totally plant based only available for NZ and AUS residents only.

Figure 4.2 sample view of constructed corpora from collected spam emails from 2021-01 from 2022-04

4.3.2 Structure of the developed python program to train the model and to detect spam email

To build the model I have chosen the TensorFlow framework which is developed and maintained by Google. The TensorFlow framework is using Long Short Term Memory neural network algorithm which Is classified as a recurrent neural network. The reason for using LSTM-based algorithm for text classification is that traditional neural networks endure short-term memory loss and are hard to find the relevant information from old memory (Shekhar, 2021). On the other hand, LSTM provides meaning in a full sentence when it comes to words that have several meanings. For example, “Cricket, Fly”. These kinds of words can provide meaning based on where and how it is used.

Step 1: Text Sanitisation

First, all the raw text contains special characters and wildcard characters, therefore before you process the raw email containing text messages it needs to be sanitised. Commonly this process treats punctuations or string punctuations. In addition, in this pre-processing stage, it will remove numbers, and any hyperlinks, white spaces and replace the new lines. The following image shows the spam email body before treatment and after removing the punctuation.

Before treated	After Treated
Take part in our marketing survey and Get $90 promo reward	take part in our marketing survey and get promo reward
Take action TODAY - Up to 50% off on Premium subscriptions. Give it a go! It's time to fall in love. You deserve it,	take action today up to off on premium subscriptions give it a go its time to fall in love you deserve it
Play the Pets game to meet new people!	play the pets game to meet new people
This is to bring to your notice that you've been dealing with scammers over the years with regards to your long standing	this is to bring to your notice that youve been dealing with scammers over the years with regards to your lo
Get your SmartMatch of today.Thousand of members get their perfect match on this site!	get your smartmatch of todaythousand of members get their perfect match on this site
Post Office Notification - Automated Delivery date: 9 November 2021 Your delivery status: On hold The courier assigne	post office notification automated delivery date november your delivery status on hold the courier assig
I'm attracted to guys with long hair. Leather jacket is a bonus. send a pic	im attracted to guys with long hair leather jacket is a bonus send a pic
More People To Show Interest In á´„Ê€á´™á´›0 In 2022	more people to show interest in á´„Ê€á´™á´› in
â,¬ 3000 welcome Bonus + 175 Free Spin Prices and products in effect November 12-17, 2021, unless otherwise stated. F	â,¬ welcome bonus free spin prices and products in effect november unless otherwise stated prices are
you have (1) package waiting for delivery. Use your code to track it and receive it. Schedule your delivery and subscribe	you have package waiting for delivery use your code to track it and receive it schedule your delivery and su
Well done Pubudu! Your email address was selected! Please make sure you claim your chance to receive this $2,000 PA	well done pubudu your email address was selected please make sure you claim your chance to receive this
Fedex#44259, We've got a new message for you. We have sent you a message with the required information. Have tro	fedex weve got a new message for you we have sent you a message with the required information have tr
WhatsApp Missed voice message. Details Date: Dec 23 3:40 Duration: 09 seconds Play	whatsapp missed voice message details date dec duration seconds play
Dear Pubudu Gayan Delivery status: Cancelled Last delivery attempt: 45 minutes ago Current package location: Logistics	dear pubudu gayan delivery status cancelled last delivery attempt minutes ago current package location lo
Hi PUBUDGYAN, We are sure you appreciate our protection against dangerous viruses, hackers and cyber criminals. Unf	hi pubudgyan we are sure you appreciate our protection against dangerous viruses hackers and cyber crimi
African â€œTwistâ€ Technique..Adds-6-Inches	african â€œetwistâ€ techniqueaddsinches
The Amazing Keto Supplement! Melt Fat Fast without diet or exercise. Totally plant based only available for NZ and AU	the amazing keto supplement melt fat fast without diet or exercise totally plant based only available for nz
Track your package (Expiry Date : 48 hours) Get your supended packckage in next 24 hours, click the link here https://en	track your package expiry date hours get your supended packckage in next hours click the link here
DEADLINE: 1 week left DONATE $25 DONATE $50 DONATE $100 DONATE $150 Donate other amount	deadline week left donate donate donate donate donate other amount
5 Dec GOLDEN E-TICKETSðŸ'¥$2.50 Yomie's Rice x Yogurt add to your cart now! ðŸ'¥	dec golden etickets ðŸ'¥ yomies rice x yogurt add to your cart now ðŸ'¥
A newly single experienced woman would like to meet a younger man preferably early/mid-20s. Age is just a number, i	a newly single experienced woman would like to meet a younger man preferably earlymids age is just a nu

Figure 4.3 Text before treating and after treating.

Step 2: Tokenizing and Padding

Secondly, the treated text needs to be tokenised. In general, tokenization means splitting the long sentence into words or smaller units. For the tokenization process, I have used the "Keras" library as it supports TensorFlow and is also based on a neural network in python. The Keras library is easy to use and implement when it needs to do the tokenization process.

Padding is a necessary step when it comes to NLP, because almost every neural network needs the sentences to be equal. As we know in real life sentence lengths are usually different from each other (Shrestha, 2020). Therefore, the process called Padding comes into use, so every sentence makes an equal length. As I am using RNN it helps to understand the sentence from the beginning therefore in this instance I have employed pre-padding.

Step 3: Creating the model and training the model.

In the fourth step, I need to build the model and train it for the text analysis. In Figure 4.4 the model summary is shown.

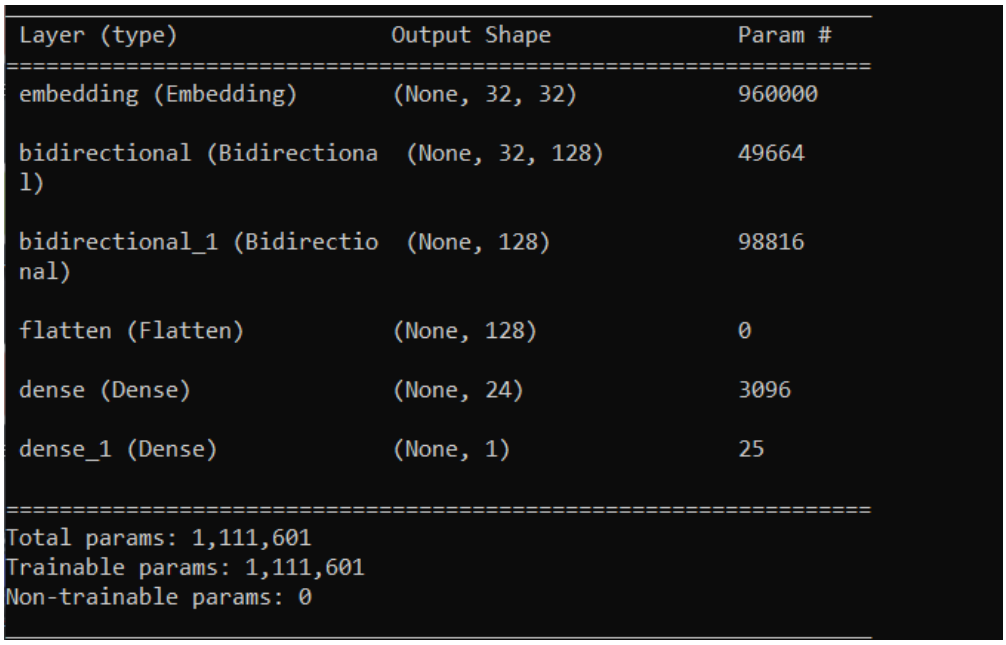


Figure 4.4 Model Summary

When there is a small amount of training (spam emails) the model is not accurate and model accuracy shows a flat line against the accuracy. However, eventually, I added more training data and model accuracy gradually increased. Figures 4.5, 4.6, and 4.7

illustrate how the model accuracy has increased when you increase the training data.

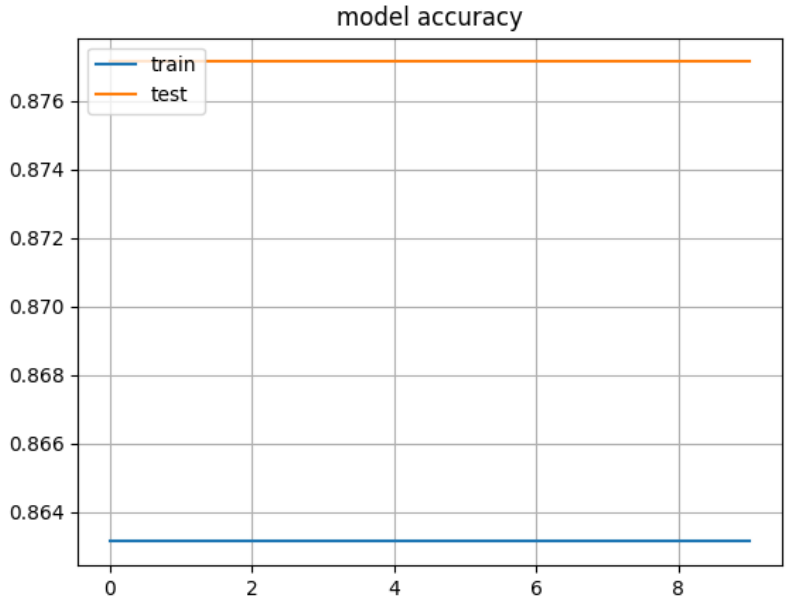


Figure 4.5 Model accuracy less than 500 emails

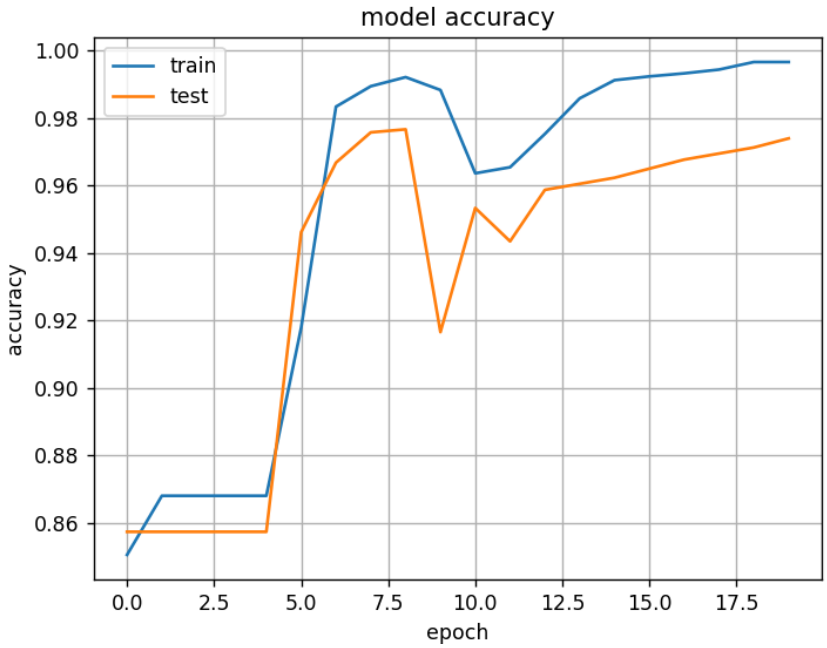


Figure 4.6 Model accuracy less than 2500 emails

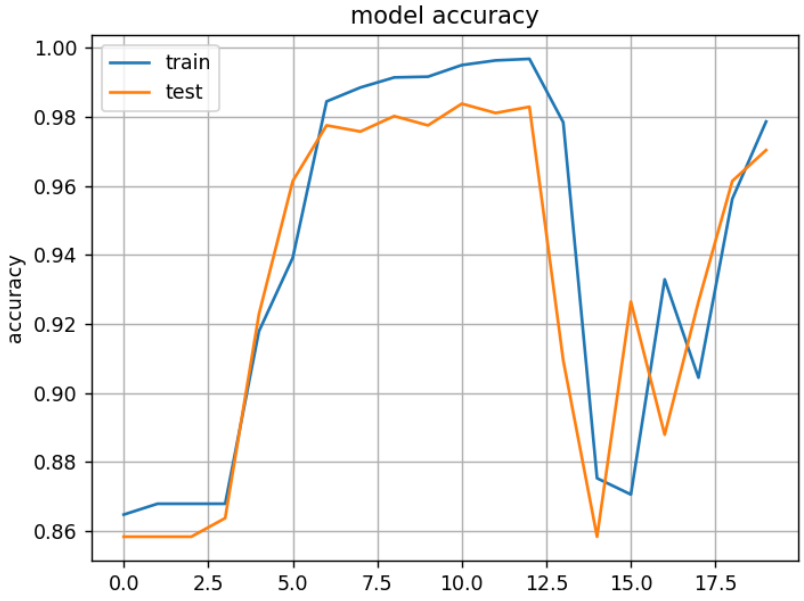


Figure 4.7 Model accuracy less than 3000 emails

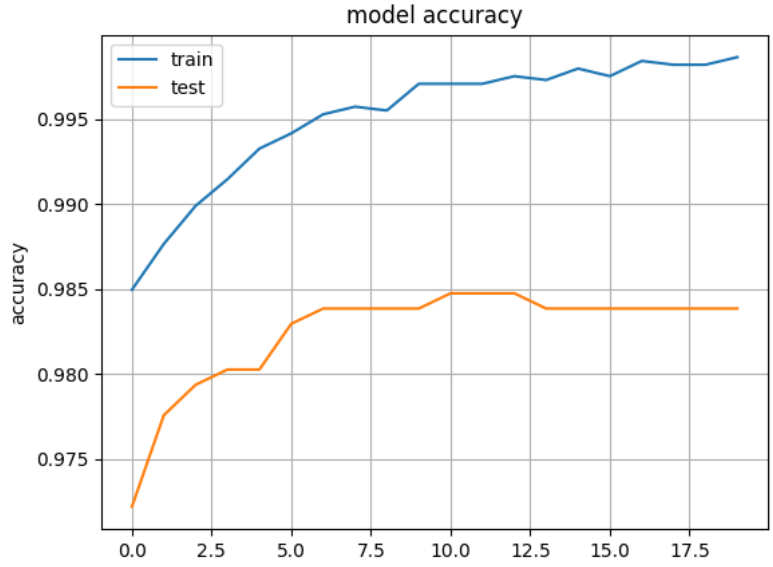


Figure 4.8 Model accuracy with more than 6500 emails

After training the model it can be saved for future processes or analyses.

Step 4: Confusion matrix Analysis

According to Bhandari (2020), a confusion matrix estimates the model performance. By using the confusion matrix, it can understand how the model's accuracy plus what kind of errors it produces (Bhandari, 2020).

Figure 4.9 shows the confusion matrix that the model created in the previous step.

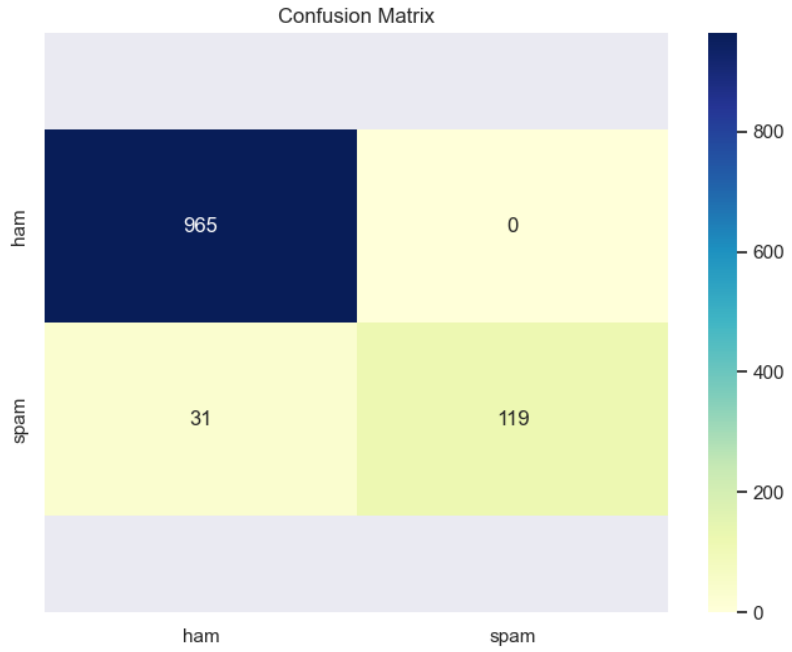


Figure 4.9 Confusion matrix

In the General confusion matrix, the True positive (TP) section represents the predicted values that match the actual model values (Bhandari, 2020). Furthermore, True Negative (TN) value represents as same as TP (the forecasted values represent the real values) but the actual value is negative, and the model forecasted value is negative too. The third section of the confusion matrix is represented as the False Positive (FP), which means the forecasted value is wrongly predicted and the real value is negative although the model forecasted a positive value (Bhandari, 2020). In addition, these values are known as Type 1 errors. Finally, False Negative (FN) values represent the Type 2 errors in the model. FN presents a forecasted value that is falsely predicted, and the real value is positive however in this case the model forecasts the negative values (Bhandari, 2020).

Though, in this research TP represent the total number of Ham messages correctly detected as ham messages (965). In addition, 119 counted as True Negative, which means spam messages were detected as spam messages. On the other hand, 31 messages have been falsely predicted, which means detected spam messages are detected as ham messages. Most importantly in this research, no false negative messages have been detected.

Hagiwara (2021) asserts that precision represents the percentage value of the correct result from the entire result set, in other words, how many predicted items are

correct in the actual result set. The following equations are reproduced from section 3.1.2, equations (1) to (4) to compute the relevant performance values from the experimental data.

$$\text{Precision} = TP/(TP+FP)$$

(Hagiwara, 2021)

Therefore, according to the confusion matrix shown in figure 4.9 precision is 1.

$$\text{Precision} = 965/(965 + 0)$$

Recall indicates how many actual positive values were predicted correctly in our NLP model (Bhandari, 2020). Ultimately, recall value calculates as follows.

$$\text{Recall (a)} = TP / (TP+FN)$$

(Bhandari, 2020)

As same precision according to the confusion matrix in this research model is 0.9688

$$\text{Recall (b)} = 965/ (965+31)$$

By combining the recall and precision values we can find the model accuracy. The reason is the model contains false values and true values as well. In building the model and using the build model for NLP this false and true both values are equally important. This is because identifying false negatives (FN) can reduce them in the model and to increase the accuracy of the model. To calculate the model's accuracy, it needs to calculate the F1 score because it represents the imbalance of item distribution within the model (Hagiwara, 2021).

F1 score = 2 x precision x recall / (precision + recall)

$$= 2 \text{ x } a \text{ x } b / (a+b)$$

$$= 2 \text{ x } 1 \text{ x } .9688 / 1.9688$$

$$= 0.9841$$

Masato Hagiwara (2021) says in his book that if the F1 score is equal to 1 it means the model is perfect and if it scores 0, that means the model is incorrect. Hence, according to the above calculation, the F1 score in this research model is almost correct.

Step 5: Using the model to detect spam emails.

In the last step, uses the model to check whether email text is spam or ham. For this research, I have used saved email text and scanned it using the model. However, the following Figure 10 will illustrate the spam percentage of email content of collected test emails.

Before treated	After Treated	Spam %
Take part in our marketing survey and Get $90 promo reward	take part in our marketing survey and get promo reward	15%
Take action TODAY - Up to 50% off on Premium subscriptions. Give it a go! It's t	take action today up to off on premium subscriptions give i	0.00%
Play the Pets game to meet new people!	play the pets game to meet new people	99.56%
This is to bring to your notice that you've been dealing with scammers over th	this is to bring to your notice that youve been dealing with s	0.06%
Get your SmartMatch of today.Thousand of members get their perfect match c	get your smartmatch of todaythousand of members get thei	0.01%
Post Office Notification - Automated Delivery date: 9 November 2021 Your de	post office notification automated delivery date novembe	0.00%
I'm attracted to guys with long hair. Leather jacket is a bonus. send a pic	im attracted to guys with long hair leather jacket is a bonus s	99.76%
More People To Show Interest In á´„Ê€Êá´™á´›0 In 2022	more people to show interest in á´„Ê€Êá´™á´› in	Error
â‚¬ 3000 welcome Bonus + 175 Free Spin Prices and products in effect Novemb	â‚¬ welcome bonus free spin prices and products in effect	Error
you have (1) package waiting for delivery. Use your code to track it and receive	you have package waiting for delivery use your code to trac	0.01%
Well done Pubudu! Your email address was selected! Please make sure you cl	well done pubudu your email address was selected please	99.93%
Fedex#44259, We've got a new message for you. We have sent you a message	fedex weve got a new message for you we have sent you a	99.89%
WhatsApp Missed voice message. Details Date: Dec 23 3:40 Duration: 09 secon	whatsapp missed voice message details date dec duration	99.05%
Dear Pubudu Gayan Delivery status: Cancelled Last delivery attempt: 45 minut	dear pubudu gayan delivery status cancelled last delivery at	0.00%
Hi PUBUDGYAN, We are sure you appreciate our protection against dangerous	hi pubudgyan we are sure you appreciate our protection aga	0.01%
African â€œTwistâ€ Technique..Adds-6-Inches	african â€œtwistâ€ techniqueaddsinches	Error
The Amazing Keto Supplement! Melt Fat Fast without diet or exercise. Totally	the amazing keto supplement melt fat fast without diet or e	99.75%
Track your package (Expiry Date : 48 hours) Get your supended packckage in ne	track your package expiry date hours get your supended pa	5.95%
DEADLINE: 1 week left DONATE $25 DONATE $50 DONATE $100 DONATE $150 D	deadline week left donate donate donate donate donate	99.89%
5 Dec GOLDEN E-TICKETSðŸ"¥$2.50 Yomie's Rice x Yogurt add to your cart now!	dec golden eticketsðŸ"¥ yomies rice x yogurt add to your ca	Error
A newly single experienced woman would like to meet a younger man prefer	a newly single experienced woman would like to meet a yo	99.65%

Figure 4.10 Spam test result

Figure 4.10 is showing that some spam emails come with the email name (for test purposes I have setup emails pubudugayan###@####.com). Therefore, when a spam email comes with the initial greeting the trained model does not detect those emails as spam.

4.4 ANALYSIS OF SPAM EMAIL AND SOCIAL ENGINEERING

In this research, it analyses how users' behaviour on the Internet can influence getting spam emails. In general, users get various kinds of spam email attacks. According to the imperva website (2022), social engineering attacks use psychological influence and trick business email users to click a malicious link or download ransomware. Furthermore, Vergelis stated on the Kaspersky website that spammers are using Google services and their content. Hence, there is no question that spammers use data collectors' data and services to mislead email users.

Consequently, this section analyses the inverse relationship between users' Internet behaviour and inbound spam emails.

The following graphs illustrate the spam email inbound count and user activity on the Internet. This graph represents all the spam emails collected from 5 devices in the test environment. In addition, some other spam emails are filtered out. For example, in the period of 12th January to 27th January spam emails not related to Adult entertainment topics are filtered out from the Figure 4.11 graph.

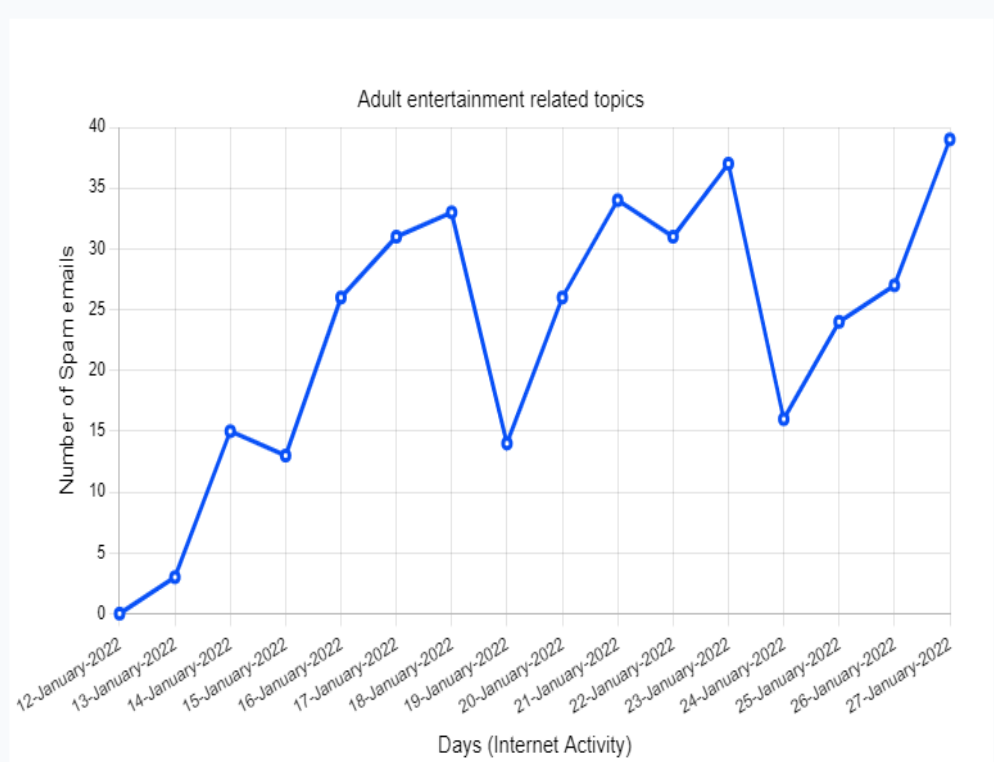


Figure 4.11 Spam email count when user search adult entertainment related topics in the Internet

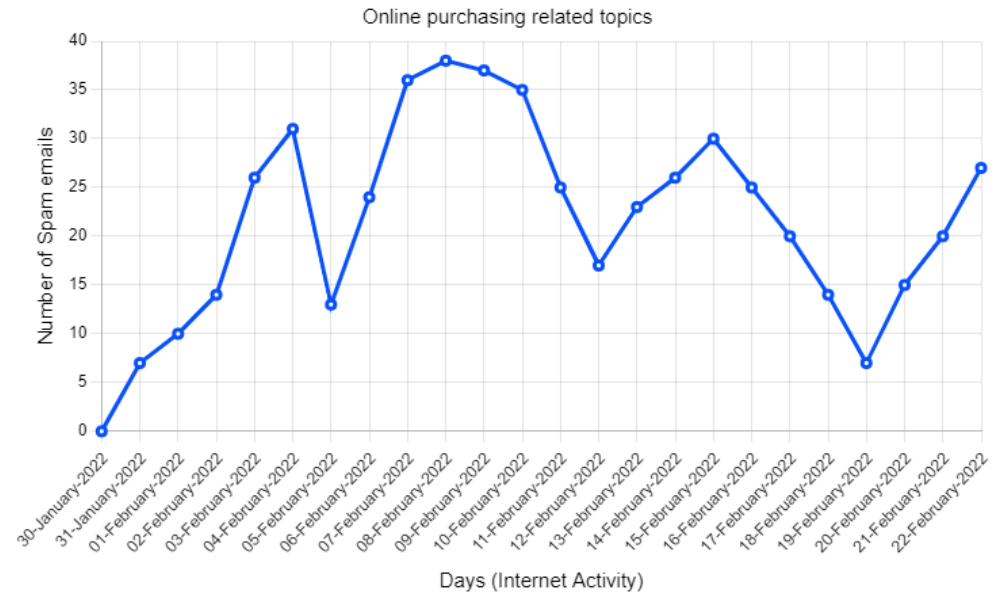


Figure 4.12 Spam email count when user search online purchasing related topics in the Internet

Figure 4.13 Spam email count when user search cryptocurrency and financial related topics in the Internet

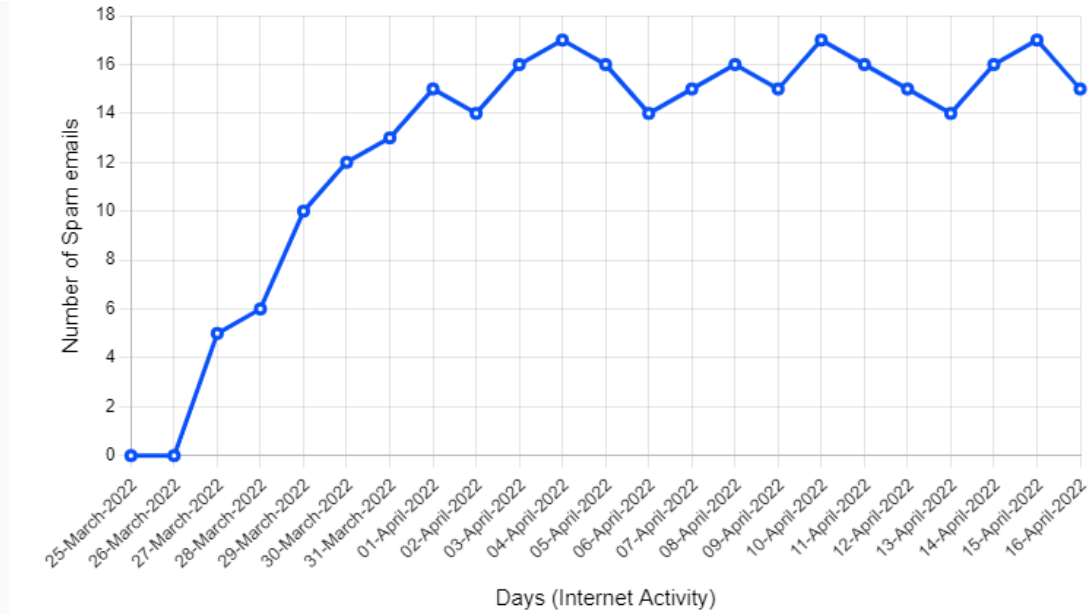


Figure 4.14 Spam email count when user search Medicine and health issue related topics in the Internet

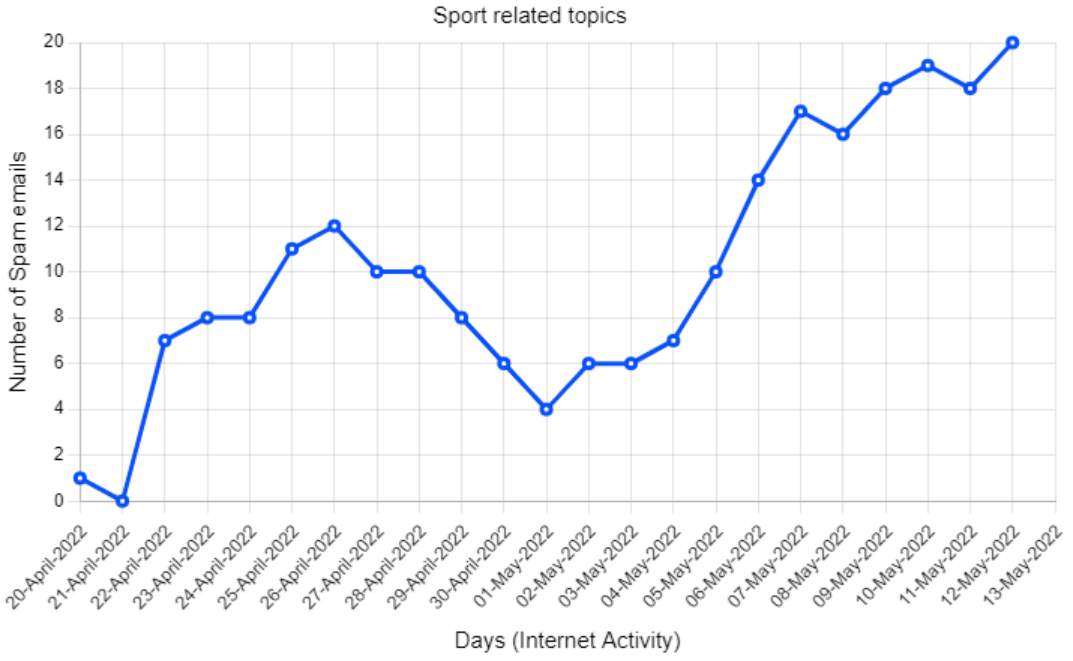
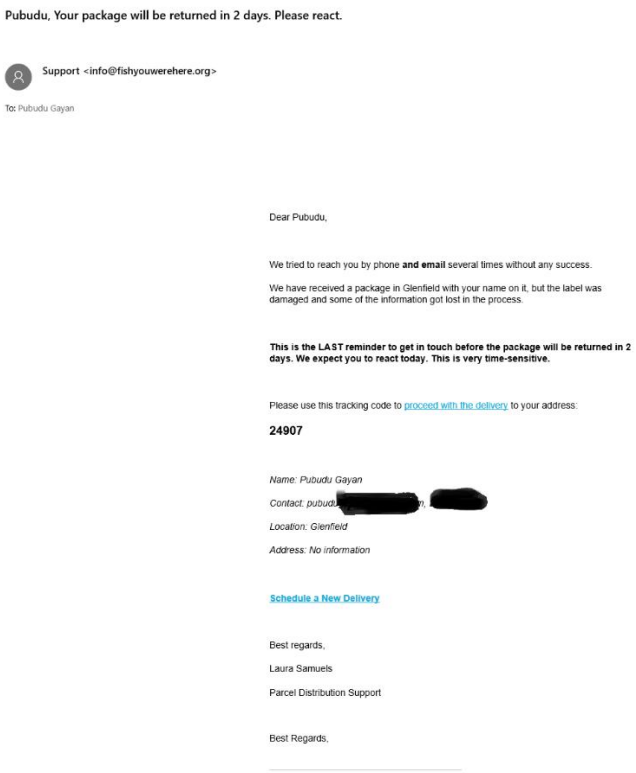


Figure 4.15 Spam email count when user search sports related topics in the Internet

In figure 4.11, 4.12, and 4.13, peak values occur when the search topic is changed to a totally different topic. For an example figure 4.12 represent the spam emails related to online purchasing topics. Plus, in that graph, all three peak points represent when I searched for a different topic unrelated to online purchasing two days before. Therefore, it clearly indicates that when the user changes the search engine topics

spam emails also follow the same pattern. However, this behaviour is not reflected on medicine and sport related topics.

Secondly, I analysed spams emails headers to identify how spammers craft spam emails in the modern day. The following email has not been detected as spam email by the trained NLP model. Then I extracted email headers and checked their domains to verify are they legitimate. Figure 4.16 shows what spam email looks like in an inbox.


Pubudu, Your package will be returned in 2 days. Please react.

Support <info@fishyouwerehere.org>

To: Pubudu Gayan

Dear Pubudu,

We tried to reach you by phone **and email** several times without any success.

We have received a package in Glenfield with your name on it, but the label was damaged and some of the information got lost in the process.

This is the LAST reminder to get in touch before the package will be returned in 2 days. We expect you to react today. This is very time-sensitive.

Please use this tracking code to proceed with the delivery to your address:

24907

Name: Pubudu Gayan

Contact: pubudu[redacted]

Location: Glenfield

Address: No information

Schedule a New Delivery

Best regards,

Laura Samuels

Parcel Distribution Support

Best Regards,

Figure 4.16 Well-crafted spam email.

It looks as a legitimate email and probably even a well-trained business email user might also trick click the provided links. Other issue is that this email came when I performed online purchasing activities in the Internet. By taking another step, I have checked the email's header details by using two recognized tools which are Mxtoolbox (https://mxtoolbox.com/) and Google Apps Toolbox (https://toolbox.googleapps.com/apps/messageheader/)

According to the MXToolbox, spam email DKIM value (DomainKeys Identified Mail) has an issue, but Google App toolbox said spam email DKIM values are fine (Figure 4.17 and Figure 4.18 illustrate the results). As this result is conflicted, I have checked the domain health with MXToolbox, and according to it,

the spam email domain (fishyouwerehere.org) has been blacklisted (Figure 4.19 present the MXToolbox results about spam email domain). Therefore, this is good evidence that even email users who have technical knowledge also cannot depend on one tool to validate whether their email is a legitimate email or not.

Delivery Information

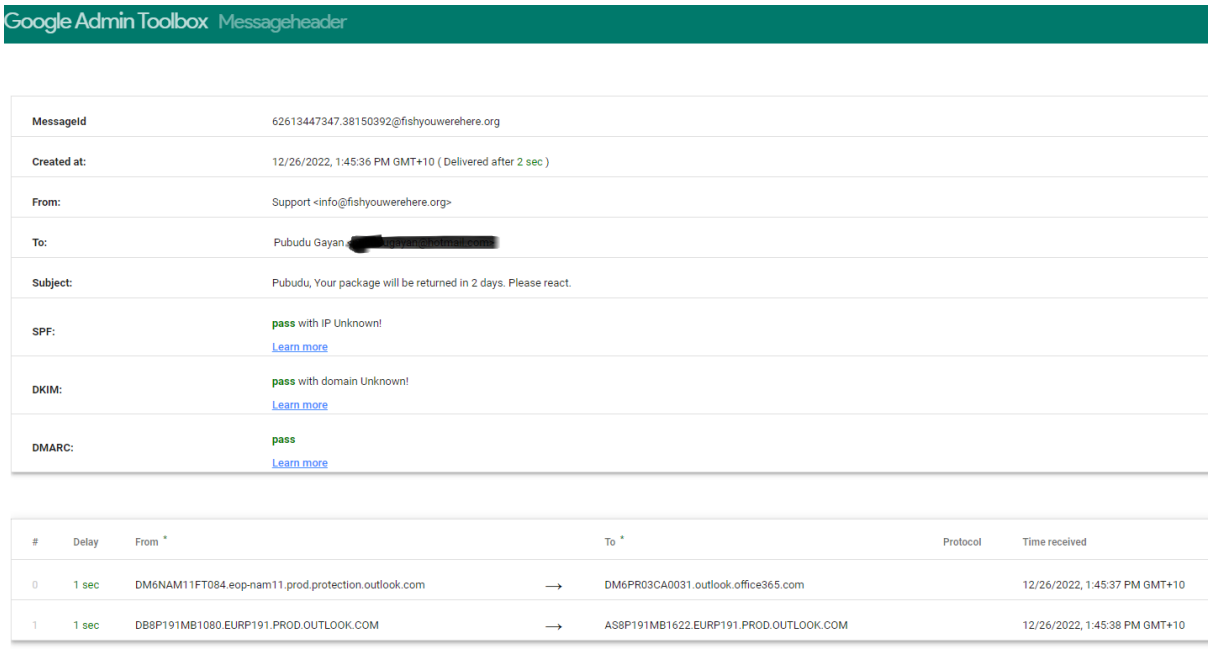
- DMARC Compliant
  - SPF Alignment
  - SPF Authenticated
  - DKIM Alignment
  - DKIM Authenticated

Relay Information

Received Delay:	2 seconds

Hop	Delay	From	By	With	Time (UTC)	Blacklist
1	*	mxptr1.fishyouwerehere.org 46.105.36.187	DM6NAM11FT084.mail.protection.outlook.com 10.13.172.132	Microsoft SMTP Server (version=TLS1_2, cipher=TLS_ECDHE_RSA_WITH_AES_256_GCM_SHA384)	12/26/2022 3:45:36 AM	
2	1 Second	DM6NAM11FT084.eop-nam11.prod.protection.outlook.com 2603:10b6:5:40::cafe::21	DM6PR03CA0031.outlook.office365.com 2603:10b6:5:40::44	Microsoft SMTP Server (version=TLS1_2, cipher=TLS_ECDHE_RSA_WITH_AES_256_GCM_SHA384)	12/26/2022 3:45:37 AM	
3	1 Second	DM6PR03CA0031.namprd03.prod.outlook.com 2603:10b6:5:40::44	DB8P191MB1080.EURP191.PROD.OUTLOOK.COM 2603:10a6:10:160::13	Microsoft SMTP Server (version=TLS1_2, cipher=TLS_ECDHE_RSA_WITH_AES_256_GCM_SHA384)	12/26/2022 3:45:38 AM	
4	0 seconds	DB8P191MB1080.EURP191.PROD.OUTLOOK.COM 2603:10a6:10:160::13	AS8P191MB1622.EURP191.PROD.OUTLOOK.COM	HTTPS	12/26/2022 3:45:38 AM	

Figure 4.17 MXToolBox result about spam email header

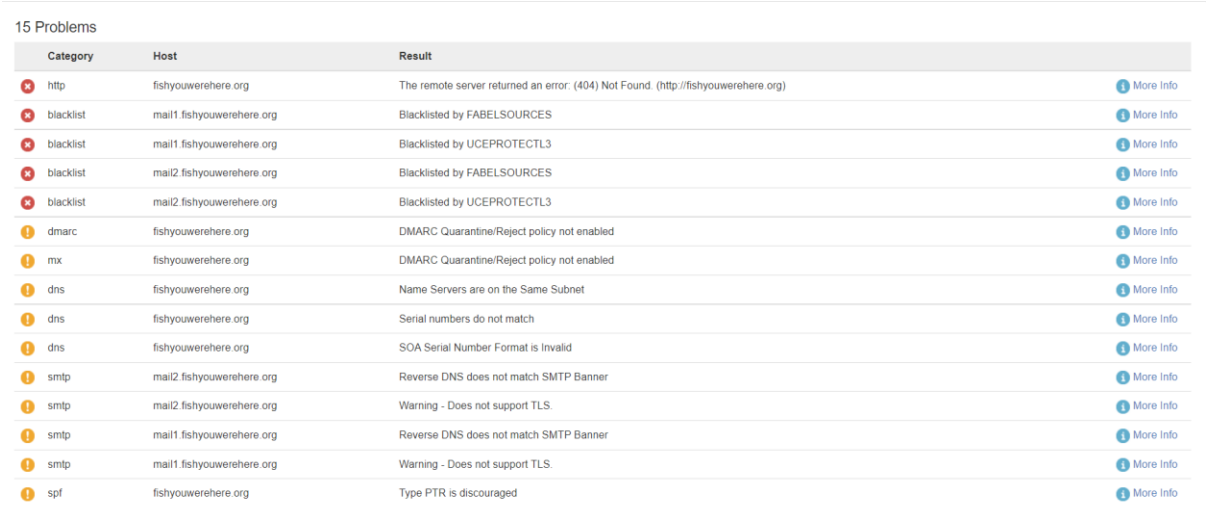


Google Admin Toolbox Messageheader

MessageId	62613447347.38150392@fishyouwerehere.org
Created at:	12/26/2022, 1:45:36 PM GMT+10 (Delivered after 2 sec)
From:	Support <info@fishyouwerehere.org>
To:	Pubudu Gayan [redacted]
Subject:	Pubudu, Your package will be returned in 2 days. Please react.
SPF:	**pass** with IP Unknown! Learn more
DKIM:	**pass** with domain Unknown! Learn more
DMARC:	**pass** Learn more

#	Delay	From *		To *	Protocol	Time received
0	1 sec	DM6NAM11FT084.eop-nam11.prod.protection.outlook.com	→	DM6PR03CA0031.outlook.office365.com		12/26/2022, 1:45:37 PM GMT+10
1	1 sec	DB8P191MB1080.EURP191.PROD.OUTLOOK.COM	→	AS8P191MB1622.EURP191.PROD.OUTLOOK.COM		12/26/2022, 1:45:38 PM GMT+10

Figure 4.18 Google Admin toolbox result about spam email header



15 Problems

Category	Host	Result	
http	fishyouwerehere.org	The remote server returned an error: (404) Not Found. (http://fishyouwerehere.org)	More Info
blacklist	mail1.fishyouwerehere.org	Blacklisted by FABELSOURCES	More Info
blacklist	mail1.fishyouwerehere.org	Blacklisted by UCEPROTECTL3	More Info
blacklist	mail2.fishyouwerehere.org	Blacklisted by FABELSOURCES	More Info
blacklist	mail2.fishyouwerehere.org	Blacklisted by UCEPROTECTL3	More Info
dmarc	fishyouwerehere.org	DMARC Quarantine/Reject policy not enabled	More Info
mx	fishyouwerehere.org	DMARC Quarantine/Reject policy not enabled	More Info
dns	fishyouwerehere.org	Name Servers are on the Same Subnet	More Info
dns	fishyouwerehere.org	Serial numbers do not match	More Info
dns	fishyouwerehere.org	SOA Serial Number Format is Invalid	More Info
smtp	mail2.fishyouwerehere.org	Reverse DNS does not match SMTP Banner	More Info
smtp	mail2.fishyouwerehere.org	Warning - Does not support TLS.	More Info
smtp	mail1.fishyouwerehere.org	Reverse DNS does not match SMTP Banner	More Info
smtp	mail1.fishyouwerehere.org	Warning - Does not support TLS.	More Info
spf	fishyouwerehere.org	Type PTR is discouraged	More Info

Figure 4.19 MXToolbox result regarding spam email domain(fishyouwerehere.org).

4.5 ANALYSING PHISHING EMAILS ACCORDING TO NIST PHISHING SCALE

The National Institute of Standards and Technology (NIST) has introduced a scale to understand the difficulty of phishing email content. In general, it considers greeting, email content and previous conversations (The Phish Scale: NIST-Developed Method Helps IT Staff See Why Users Click on Fraudulent Emails, 2020). Therefore, to get an idea about how the actual users' reactions regarding these phishing emails, conducted a small survey by inviting ten actual business email users (These users are well-trained regarding cyber security from their workplaces).

For that survey they have chosen 5 phishing emails that were collected from the phase 2 spam email collection phase. Figures 4.20 to 4.24 shows how they look and feel. The survey results show in the following table (Table 4.2)

	Attack Load	Difficulty Level Number of Votes			
		Very Difficult	Difficult	Moderate	Easy
Figure 4.20	Link	7	2	0	1
Figure 4.21	Link	0	0	1	9
Figure 4.22	Link	0	2	4	4
Figure 4.23	Link	0	4	2	4
Figure 4.24	Link	0	0	0	10

Table 4.2 Phishing email difficulty level according to ten business email users

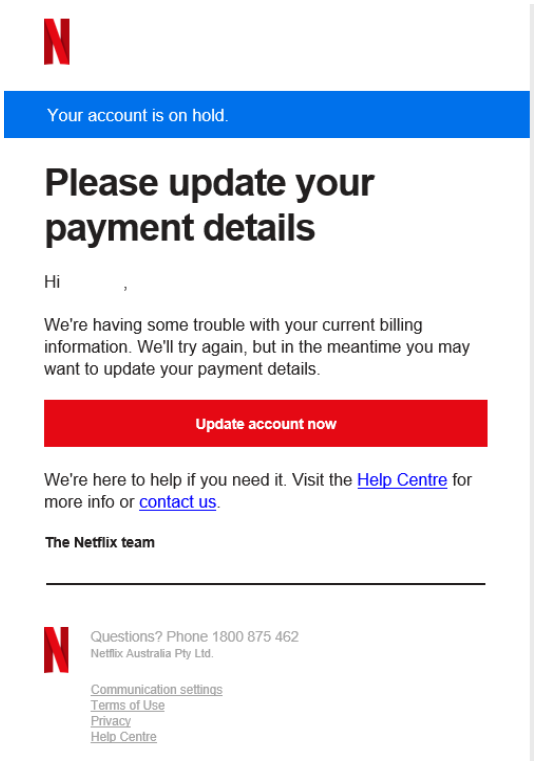


Figure 4.20 Netflix account update

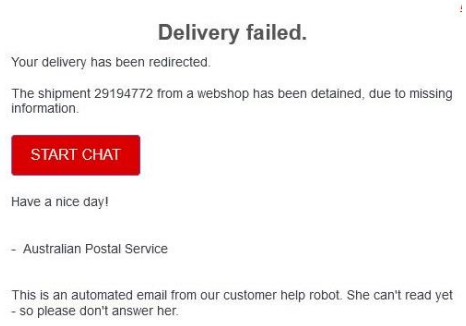
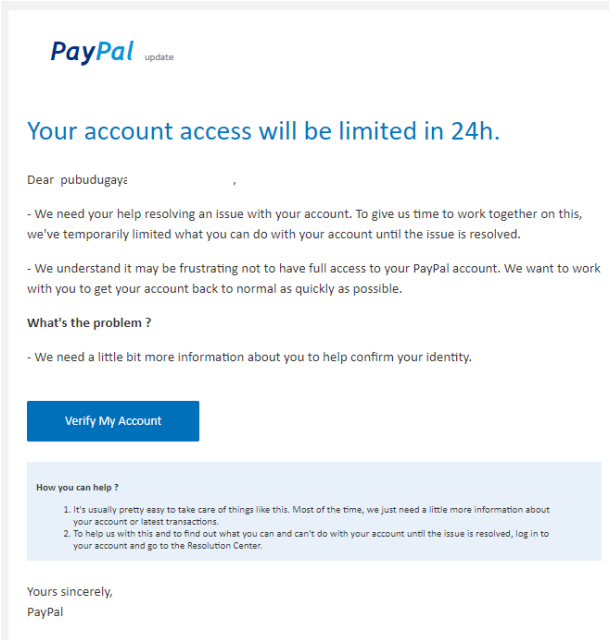


Figure 4.21 Australia Post Delivery failed



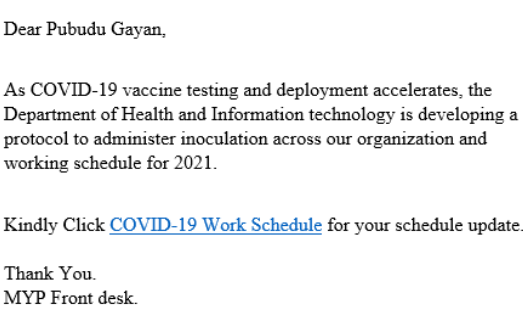


Figure 4.22 Paypal account verification

Figure 4.23 Covid vaccination and work schedule

Dear Pubudu,

We tried to reach you by phone **and email** several times without any success.

We have received a package in Glenfield with your name on it, but the label was damaged and some of the information got lost in the process.

This is the LAST reminder to get in touch before the package will be returned in 2 days. We expect you to react today. This is very time-sensitive.

Please use this tracking code to proceed with the delivery to your address:

24907

Name: Pubudu Gayan

Contact: pubudugaya

Location: Glenfield

Address: No information

Schedule a New Delivery

Best regards,

Laura Samuels

Parcel Distribution Support

Best Regards,

Figure 4.24 Reschedule delivery.

4.6 SUMMARY OF ANALYSIS

In the literature review it has discussed various ways that cyber criminals are using to compromise the business email accounts. Therefore, in this section it has tested a trained NLP model to prevent spam emails, but test results showed (Figure 4.10) when spam email has a greeting and email recipient name, this well-trained NLP model also cannot recognise the spam email correctly. Even further when manually analysed the spam email headers it returns incorrect results from market leader tools that are recognised for email security.

In general, it shows there is a requirement for a new security framework to protect business emails from spam emails by considering users' behaviour within the Internet. This is because in this research test results clearly indicate spammers are getting user behaviour data via search engine data collectors and according to that data, spammers shape up the spam emails to trick legitimate users into compromising business email systems.

Secondly, this research survey showed if it well-crafted spam email is received, it is a difficult job to identify the spam emails from the legitimate emails for the email users. Furthermore, there are lots of new tools coming to the technology market based on AI that generate text look like legitimate emails (like ChatGPT). This is a challenging task for security implementers because the AI generates text to exactly look like legitimate conversation.

4.7 CONCLUSION

Chapter 4 was dedicated to reporting building the NLP model to find spam emails from legitimate emails. However, even with the 99% accurate NLP model also fails when spam emails have proper greetings with the recipient's appropriate name. Furthermore, in this chapter 4, the relationship between users' Internet behaviour and inbound spam emails is reported by testing actions and the data analysis. Spam email creators are using users' Internet searches from third parties and carving well-formatted spam emails to trick business email users into compromising business emails. Additionally, even by inspecting the email header, it is hard to identify the spam email because different email header analysis tools give different results. Therefore, in chapter 5 I will discuss this chapter's findings and examine the research question plus present the new email security framework that can minimize the spam email attack impacts.

Chapter 5

Discussion of Findings

5.0 INTRODUCTION

In the literature review it has discussed various types of email attacks that spammers and cyber criminals are using to breach business email security. Furthermore, there are many mail protocols and security products available. However, these security barriers can be compromised by email users' mistaken actions. Section 2.2 illustrates how business email can be compromised in the modern cyber security domain. It is identified that human factors are the most crucial and important factor when it comes to breaching business email security systems. Section 3.1 has been dedicated to reviewing four similar studies, and that review helped to build the research methods and identify the technologies and techniques that need to be used to answer the research question. The research findings and the behaviour of the built NLP model was presented in chapter 4. Additionally, chapter 4 has shown that users' behaviour on the Internet is influencing incoming spam emails.

The main idea of chapter 5 is to discuss chapter 4 findings and develop the answer to the research question by using the results in chapter 4 and compare them with the literature views in chapter2. Chapter 5 contains four sections. The first section 5.1 will focus on assessing the sub-questions according to the research results. This research has two sub-questions to evaluate in order to find the answer to the main research question. Therefore, the two sub-questions are SQ 1 "What type of attacks and methods are hackers using for compromising business email systems?" and SQ2 "How do hackers use social engineering techniques to influence business email users to open unsafe emails?". Section 5.2 reviews the research hypothesis which is "The new Corpora will add at least 2000 category records but might still miss new spam or phishing email attacks". The main research question will be addressed in section 5.3, and section 5.4 presents the proposed business email security framework. Lastly, Section 5.5 concludes chapter 5.

5.1 SUB QUESTIONS

This section will evaluate the sub questions that were formed in section 3.2.2. These two sub question answers will help to build the answer to the main research question and help to build the new business email security framework.

5.1.1 Sub Question 1

SQ 1: What type of attacks and methods are hackers using to compromise business email systems?

This sub question was formed to get an idea of what type of attacks and methods spammers use to compromise the business email system. Figures 4.11, 4.12 and 4.13 have shown that inbound spam emails have a relationship with user behaviour or else what a user searches on the Internet. Furthermore, these findings have been confirmed by Vergelis (2019) publication on Kaspersky web blog. Additionally, perpetrators are now well-educated about how spam filters work. The best example is figure 4.16 which shows how they crafted the email text message and technical details as well. Even technical people find it hard to identify whether a spam email is spam or not by checking its DKIM value when Google says it is not spam and MXToolbox says that the email domain is blacklisted.

Secondly, when this email contains text messages scanned with the trained NLP model, it says it is a legitimate email by showing 99.78% accuracy. Therefore, even these well-shaped emails cannot be detected by NLP as well. However, in contrast, in the survey trained business email users have identified this email as a spam email.

To conclude the sub question No1, it can state that spammers and hackers create well-defined spam emails that cannot be recognised by NLP plus trained business email users as well.

5.1.2 Sub Question 2

SQ 2: How do hackers use social engineering techniques to influence business email users to open unsafe emails?

This sub question is designed to understand how perpetrators are using social engineering techniques to breach business email security. In this research, I have designed the test environment to monitor the users' Internet behaviour and record the spam email inbounds according to what they searched on the Internet. As we can see in Figures 4.20 to 4.24 almost all spam emails are well-designed and look like legitimate emails.

When it is closely examined Figures 4.11, 4.12, 4.13 graphs show that when the users change their search topics on the Internet, the spam email count gets low within less than two days on average. This is a fundamental issue that, in general security protocols developers or cyber security developers are missing from their security frameworks. In addition, as mentioned in sub question 1, business email users who are trained in spam emails also get confused with well formatted phishing emails.

To conclude, the sub question gives evidence that spammers and hackers are utilising some social engineering data (users' behaviour on the Internet over third-party service providers) to send targeted spam and phishing emails.

5.2 HYPOTHESIS TESTING

This hypothesis was developed by reviewing various past literature and by studying previous similar studies. In the following table evaluates the evidence for and against the research hypotheses: "The new Corpora will add at least 2000 category records but might still miss new spam or phishing email attacks."

Argument For	Argument Against
Initial stage it used less than 2000 collected spam emails (more than 1000). However, most of the tested spam emails failed from the model which is trained by using this Corpora. The NLP model's accuracy depends on how many types of text phases and correct categories corpora have.	Even though Corpora have more than 5000 records NLP model failed when spam emails have greetings and well formatted paragraphs and endings.
Conclusion justification	

This Hypothesis is false. The NLP engine needs minimum categorised records to prepare the model. However, Figure 4.20, 4.22, 4.23, 4.24 spam email text not detected as spam email by the NLP model which is trained more than 5000 categorised text records. Therefore, it is concluded that the model needs to be train all the time by new spam email texts.

5.3 THE RESEARCH QUESTION

The main goal of this section is to determine an answer to the main question based on the findings of the sub questions and research findings in chapter 4. The research question is: “What patterns are detected in email attacks on business systems?”. This has shaped and directed the entire research project. The primary scope of this research is to find out how business email can be compromised via spam emails and find out the patterns between users’ Internet behaviours. In a few years, back spam emails can be identified by grammar mistakes and by investigating the originating IP address or else by looking at the subject details. However, with the technology revelation, spammers get away from those mistakes and with the cloud service providers they managed to hide their hosting IP addresses.

The best way to prevent spam attacks is to block them before they arrive in the users’ inboxes. In section 2.2.1 in the literature review has discuss the users’ actions are the most critical issue when it comes to the BEC. Therefore, the best thing to protect the security of email inboxes is to stop spam or phishing emails coming to the users’ inboxes. By using the NLP AI tool, it is expected that a well-trained model will detect spam emails. Yes the research shows that it can detect spam emails, but this NLP trained model also fails to identify spam emails when they are well written. Furthermore, it has been shown in chapter 4, section 4.5 that even users who are aware of spam and phishing emails can be misled by well written spam or phishing emails.

Section 4.4 discusses how the incoming spam email behaviour relates to the users’ Internet search contents. Most of the spam emails are related to what the users searched on the Internet in particular periods or durations. It is a well-known factor

nowadays that search engines and smart devices collect user behaviour as electronic data for marketing purposes or to improve the users' experiences. However, the dark side of this practice is that these data can be use against the users. For this research, it has used newly created emails and the test environment was only used to perform the Internet search and search about relative topics mentioned in section 3.3.1. Consequently, there is no other way that spam email creators get to know the users' behaviour in the Internet during that period other than from the users' data collectors (this can be a search engine or cookies and smart devices).

5.4 RECOMMENDATION FOR A NEW BUSINESS EMAIL SECURITY FRAMEWORK

In this section, it proposes a new security framework to improve business email users' account security. Basically, this framework only considers individual users' email accounts and is not designed to protect the organisation's domain email security. In general, this framework has three main components those are, User behaviour model, the Forecasting model, and the Text analysing model (NLP model). This research has confirmed spam email creators are monitoring user behaviour on the Internet and use the information they create spam emails, to target individuals. They are good enough to fool a NLP based spam email categoriser. Furthermore, when designing this framework, it considers the survey results that are shown in table 4.2. The proposed framework is in figure 5.1. The next section describes the individual modules and their mechanisms.

In General, these three modules are working together or else can work individually. The reason for that is sharing knowledge between these three models is helping to identify more false emails. However, there is no restriction in this framework so that individual modules can work independently.

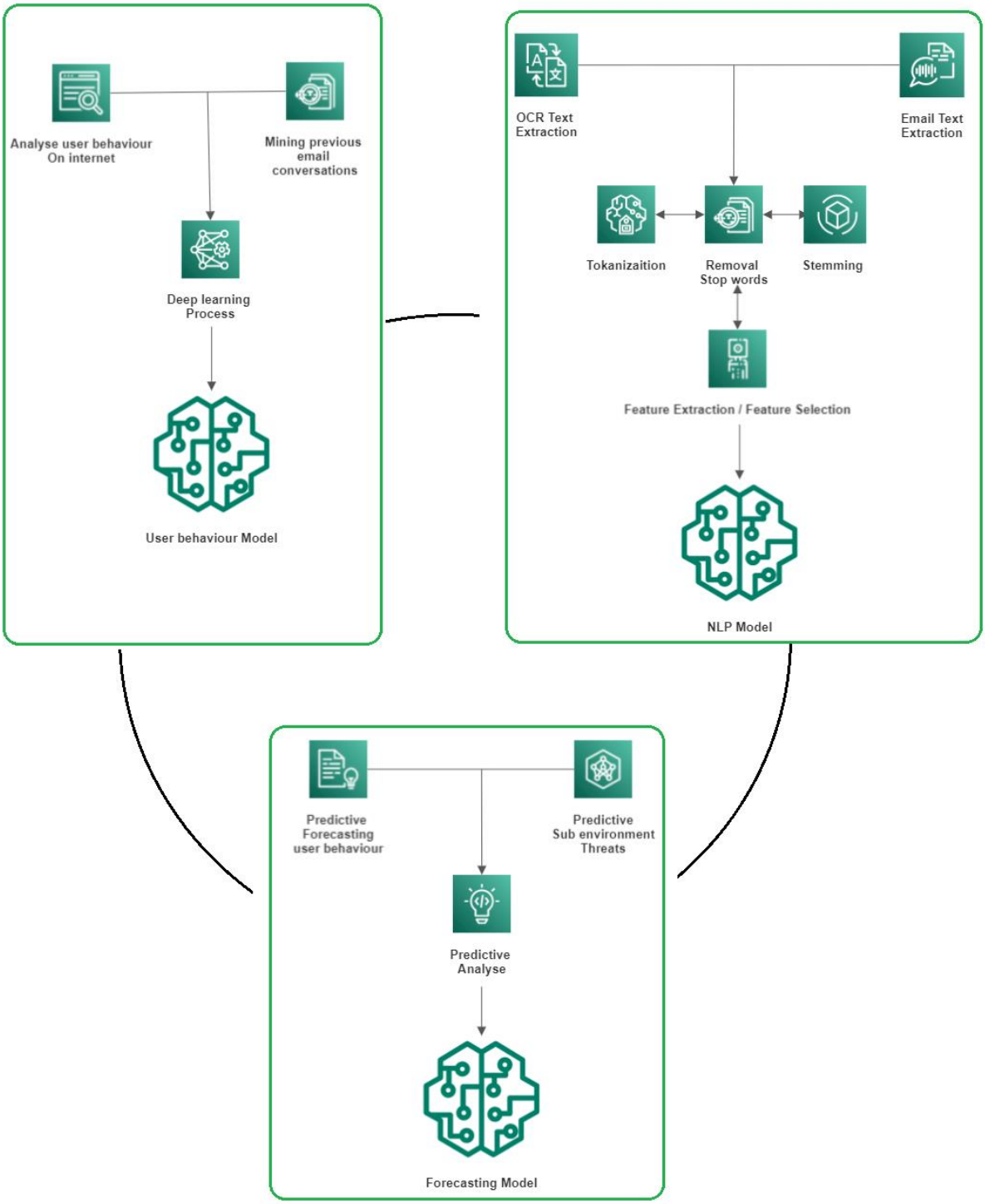


Figure 5.1 Proposed security framework for business users email accounts

5.4.1 User behaviour module

The principal behind this module is for analysing the user behaviour, this can be on the Internet and the previous email conversations.

- **Analyse User behaviour on Internet:** The main responsibility in this component is monitor the user behaviour on the Internet and analyse the search word phrases. These word phases are the most important when it categorises spam email because spam email generators are getting these word phases from the third party and manipulate business email users to execute malware (this can be a link or form or executable file).
- **Mining previous email conversations:** Analysing previous email conversations is the most important part of the behaviour analyse module. The reason for that is by analysing the previous conversation, the proposed framework can alarm the user to even well-crafted phishing emails. For example, if a client asks to pay a different account than the previous account, then this module can be trained to extract all the information and alarm the users or other relevant parties to prevent human mistakes in doing the monitory transactions. By using this module, it can analyse individual email communicators writing patterns. This is a kind of DNA of writing on electronic media. For example, some users start email using Dear, Some initiate Hi, and some use Hello. Therefore, by carefully analysing if users get phishing emails from known email address systems can identify the percentage of probability and get the attention of the email users about the suspect emails.
- **Deep learning process:** combined results from the previous two modules are used for the deep learning process here. Basically, this module is responsible for training the module regularly and understanding the user behaviour patterns. This can be user behaviour on the Internet or in previous email communication.
- **User behaviour model:** This is the model that uses the scan engine to categorise spam emails or alarm the user about the probability of spam emails.

5.4.2 Text Analysing Module

Content analysing of the incoming emails is the main responsibility of this module. Text, attachments, links, pictures, calendar events are the main contents of incoming emails. By using this module, the new framework can understand incoming spam emails, and the individual module responsibilities are described below.

- **OCR Text extraction:** This module extracts the text from the images that are attached to the incoming email.
- **Email Text extraction:** This module extracts all the text from the email.
- **NLP Module:** This is the module that the scan engine uses to scan incoming emails to identify whether its spam or not. It processes Tokenization, removal stop wards, stemming, feature selection and feature extraction processes are discussed and explained in the literature view and the previous studies section as well. Therefore, in this framework, they also use the same tasks that were discussed in the previous sections.

5.4.3 Forecasting module

This is one of the major modules in this proposed security framework because this module predicts probable incoming attacks based on the other two modules.

- **Predictive forecasting user behaviour:** This sub-module is forecasting the upcoming future threats based on past user behaviours. For example, if a user searches for a product on the Internet and purchases from its online shopping website. Most predictive spam emails can expect to deliver the product and provide a new address or pay extra spam emails. Sometimes users will be trapped by this trick as they are expecting the goods via delivery. As humans can make mistakes, this module will help alarm users based on their behaviour this email can be spam or phishing email.
- **Predictive forecasting sub environment threats:** COVID pandemic pushed people to work from home. Most of the time, they used lots of third-party platforms to perform meetings and discuss other business-related matters. Via these platforms users get spam messages to email some information. A good example is SMS phishing and some phishing attacks via Microsoft teams. If users are not aware or not paying attention to this kind of attack a whole company's security can be compromised. As some third-party communications cannot be controlled by the business's main domain security framework, it needs a special module to predict and understand future attacks. Therefore, this module is responsible for predicting and forecasting third-party environment threats that can harm business email accounts.
- **Predictive Forecast analyses:** This module works as a mediator for combining the predictive analysis that comes from the above two modules.

Plus, it is also responsible to train the module that will be used to scan inbound email messages.

- **Forecasting model:** Based on forecasting techniques this model is used to scan the incoming email to identify spam emails based on a predictive AI forecasting.

5.5 CONCLUSION

This chapter has answered the main research question and supporting sub questions. Furthermore, the research hypothesis is also resolved separately by using the chapter 4 findings. The new security framework is proposed in section 5.4. To conclude this research thesis chapter 6 will discuss the unfinished works and limitations of this research. Recommendation for further research will be made.

Chapter 6

Conclusion

6.0 INTRODUCTION

Email security is facing more challenges as the hackers and attackers become more skilled with the latest technologies. As email is the mandatory communication channel in modern-day business, spammers and other false email message senders blast spam emails and unnecessary emails into business user email accounts. These actions can cause compromise of a business's entire security system and can lead to financial losses. Therefore, it cannot be ignored, or neglected and email security threats or causes that can lead BEC require address. Hence, this research has been conducted to understand business email attack patterns and potential AI NLP security help.

Chapter 6 concludes this research thesis, and the chapter will evaluate the limitations that the researcher has faced in this research. Furthermore, in section 6.2 what needs to be done for future research to improve email security and prevent business email compromise is elaborated based on what has been achieved in this project.

6.1 LIMITATIONS

Several limitations are recognised when conducting this research, as every research has specific limitations. Therefore, this section has been dedicated to addressing those limitations. Some limitations are based on knowledge, and some limitations are based on technical limitations.

NLP and ML are booming research topics today. Furthermore, NLP accuracy is mostly based on the corpora that use to train the model. There is no industry standard or guideline that researchers can follow as the base line to improve the accuracy of a NLP model. Additionally, the corpora can be variant from language to language. This research only focused on English; however, many users around the world communicate emails using different languages. Then these research results could change dramatically based on the language and the punctuation of the corpora used. Another limitation is for the research is the Bidirectional LSTM algorithm that was used based on other past research publications. There is no concrete evidence that BLSTM algorithm performs well compared to the KNN, SVM or SVM.

Present day social engineering cannot be limited to a specific source or topic, as there are various ways to harvest user behaviour details, such as social media, and mobile data (Telecom providers collect user behaviour data on their mobile phones (Stolfo, Hershkop, Wang, Nimeskern, & Hu, 2003)) plus data collection from smart devices and search engine service providers. There is no technical evidence that reports false email generators using these media to trick email users.

Thirdly, there are other communication channels used for business communication. For example, with the COVID pandemic people are forced to work remotely from their homes. Therefore, they are using MS Teams, skype, zoom, and SMS for business meetings and other business communications. However, in this research, I have not covered those areas which cannot be neglected. The reason it is not covered in this research, is the limited time period, and it is a tedious task to cover even three of them. To get the best outcome from a full test environment it needs to implement victim and attacker roles, and with as many communication channels as possible.

To build the NLP model, I have used the python programming language. For the tokenization, removing stop words, and model selection, several libraries that have already been developed were used. To accomplish the above tasks (Tokenizin and stemming), there are other libraries such as for the tokenization process that can use nltk, regular expression or else the spaCy library provides additional functionality as well. However, it would be very useful to do a technical survey before it is used for development on a particular task so comparative performances can be known. The limited timeline allocated for this research did not allow time to do a deep technical survey to select the most suitable tool for a specific task.

6.2 FUTURE RESEARCH

Email security compromising and improving the email security system are evolving topics. This is because changing technology and unpredictable human behaviour create new challenges for email security systems. Therefore, this topic needs continual research efforts with new technology and methods. For future researchers, I suggest several areas in AI. These suggestions are based on my investigation and are for someone who conducts research to improve email security based on ML and NLP.

It is well known factor that search engines and smart devices collect user behaviour data from the users' devices. However, in this research, it only covered proving the point, based on what a user searched on the Internet and according to that search, some spam emails are coming to the user's inbox. Many researchers have proved that third-party data collectors collect user behaviour data from smart devices. Hence, there is still a gap in how the search engines collect specific user behaviour, what they are collecting, and who can access those data, and in which formats. This issue is a serious concern when it comes to user privacy. The future research needs to extend my work and the work of others that prove web ads can do social engineering and influence BEC.

Finally, I would like to recommend future researchers investigate other communication channels, and how they too make security vulnerabilities for business email security systems. Currently, in Australia and New Zealand, most of mobile users are getting phishing SMS and false voice calls, and then after that, they send some emails to follow up or send some executable files to run. With the help of ML and AI future researchers can conduct research to improve users' email security. For this research area I suggest that they can combine predictive user behaviour analysis in the email, with other communication channels such as VOIP apps, SMS and search engines. The aim is to prevent false email messages entering a user inbox.

6.3 CONCLUSION

Email communication is one of the major and important communication channels for modern day businesses. Therefore, billions of email messages circulate around the world in one day. These email messages contain important and confidential data. In this research, it was found that email security can be breached even if it has an NLP trained model. Furthermore, there are limitations when the researcher conducts this type of research. There are no industrial standards to measure the number of records a corpus needs for the trained model and because of time limitations, not all security threats are covered. The threats that can come over other communications channels which email users are using on their devices need investigation. This research opens starting points for future researchers, and suggestions are made for further investigations including predictive behaviour analysis strategies.

References

(2020). *ACSC Annual Cyber Threat Report July 2019 to June 2020.* Kingston ACT: Australian Cyber Security Centre.

Banerjee, D. (2020, Apr 14). *Natural Language Processing (NLP) Simplified : A Step-by-step Guide*. Retrieved from https://datascience.foundation/: https://datascience.foundation/sciencewhitepaper/natural-language-processing-nlp-simplified-a-step-by-step-guide

Bazydło, P., Lasota, K., & Kozakiewicz, A. (2017, Nov). Botnet Fingerprinting: Anomaly Detection in SMTP Conversations. *IEEE Security & Privacy* , pp. 25-32.

Bencs´ath, B., & Vajda, I. (2007). Efficient Directory Harvest Attacks. *International Journal of Network Security*, 264-273.

Benishti, E. (2022, October 14). *forbes.com*. Retrieved from Can Machine Learning Help Prevent Business Email Compromise?: https://www.forbes.com/sites/forbestechcouncil/2022/10/14/can-machine-learning-help-prevent-business-email-compromise/?sh=1c11062b3bd3

Bhandari, A. (2020, April 17). *Everything you Should Know about Confusion Matrix for Machine Learning*. Retrieved from www.analyticsvidhya.com: https://www.analyticsvidhya.com/blog/2020/04/confusion-matrix-machine-learning/

Bikel, D., & Zitouni, I. (2012). *Multilingual Natural Language Processing Applications: From Theory to Practice.* IBM Press.

Broadhurst, R., & Trivedi, H. (2019, July 3). *Malware in Spam Email: Risks and Trends in the Australian Spam Intelligence Database*. Retrieved from https://papers.ssrn.com/: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3413442

Broadhurst, R., & Trivedi, H. (2020). *Malware in spam email: Risks and trends in the Australian Spam Intelligence Database.* Canberra ACT 2601, Australia: Australian Institute of Criminology.

Bulao, J. (2021, May 8). *The Role of AI in Cybersecurity – What Does The Future Hold?* Retrieved from https://techjury.net/: https://techjury.net/blog/ai-cybersecurity/#gref

Chivers, K. (2020, Nov 16). *What is spear phishing?* Retrieved from https://us.norton.com/: https://us.norton.com/internetsecurity-malware-what-spear-phishing.html

Columbus, L. (2020, August 12). *5 Ways Machine Learning Can Thwart Phishing Attacks*. Retrieved from forbes.com: https://www.forbes.com/sites/louiscolumbus/2020/08/12/5-ways-machine-learning-can-thwart-phishing-attacks/?sh=476ddf421035

(2017). *Consumer Intelligence Series: Protect.me.* PWC.

Cooper, E. (2021, Jan 9). *Cybersecurity expert calls for replacement technology following Tasmanian ambulance patient data leak*. Retrieved from https://www.abc.net.au/:

https://www.abc.net.au/news/2021-01-09/tasmanian-ambulance-data-breach-technology-overhaul-needed/13044780

Costales, B. (2002). Sendmail, 3rd Edition. O'Reilly Media,Inc.

Cyberattacks 2021: Phishing, Ransomware & Data Breach Statistics From the Last Year. (2022, January 18). Retrieved from https://spanning.com/: https://spanning.com/blog/cyberattacks-2021-phishing-ransomware-data-breach-statistics/

Edgaras G. (2020, March 4). *An In-Depth Guide to POP3, IMAP, and SMTP for Beginners*. Retrieved from https://www.hostinger.com/: https://www.hostinger.com//tutorials/email/pop3-imap-smtp-protocols-explained-ports

(2019). *Email: Click with Caution.* San Jose, CA: CISCO.

Fahad , E. (2020, Feb 10). *Email Architecture, Gmail two Step Verification, SMTP POP3 IMAP*. Retrieved from https://www.electroniclinic.com/: https://www.electroniclinic.com/email-architecture-gmail-two-step-verification-smtp-pop3-imap/

Felman, F. (2019, Aug 27). *Business Email Compromised (BEC) Scams Explode Under the GDPR Implementation*. Retrieved from https://www.circleid.com/: https://www.circleid.com/posts/20190827_business_email_compromised_bec_scams_explode_under_gdpr/

Five key points the public sector should consider when switching email service providers. (2023, April 20). Retrieved from https://www.themandarin.com.au/: https://www.themandarin.com.au/211958-5-key-things-to-consider-when-switching-email-service-providers-2/

Garg, P., & Girdhar, N. (2021). A Systematic Review on Spam Filtering Techniques based on Natural Language Processing Framework. *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence 2021)*, 30-35.

Gendre, A. (2018, Sep 21). *What's the Difference Between Phishing and Spear Phishing?* Retrieved from https://www.vadesecure.com/: https://www.vadesecure.com/en/blog/whats-the-difference-between-phishing-and-spear-phishing

Gibson, S., Issac, B., Zhang, L., & Jacob, S. M. (2020). Detecting Spam Email With Machine Learning Optimized With Bio-Inspired Metaheuristic Algorithms. *IEEE*, 187914 - 187932.

Greenberg, A. (2015, April 1). *Human error cited as leading contributor to breaches, study shows*. Retrieved from https://www.scmagazine.com/: https://www.scmagazine.com/home/security-news/human-error-cited-as-leading-contributor-to-breaches-study-shows/

Hagiwara, M. (2021). Real-World Natural Language Processing. In *Real-World Natural Language Processing* (p. 336). Manning.

Heinbach , C. (2020, Aug 11). *5 Types of Social Engineering Attacks*. Retrieved from https://www.datto.com/: https://www.datto.com/au/blog/5-types-of-social-engineering-attacks

Hendry, J. (2020, Jul 2). *Defence IT investment to climb to $20 billion over next decade*. Retrieved from https://www.itnews.com.au: https://www.itnews.com.au/news/defence-it-investment-to-climb-to-20-billion-over-next-decade-549965

Holtz, M. (2020, December 8). *A Beginners Guide to Email Protocols: SMTP, POP3, and IMAP*. Retrieved from https://www.liquidweb.com/: https://www.liquidweb.com/kb/a-beginners-guide-to-email-protocols-smtp-pop3-and-imap/

How Russia-linked hackers stole the Democrats' emails and destabilised Hillary Clinton's campaign. (2017, Nov 4). Retrieved from https://www.abc.net.au/: https://www.abc.net.au/news/2017-11-04/how-russians-hacked-democrats-and-clinton-campaign-emails/9118834

How scammers abuse Google Search's open redirect feature. (2020, May 15). Retrieved from https://nakedsecurity.sophos.com/: https://nakedsecurity.sophos.com/2020/05/15/how-scammers-abuse-google-searchs-open-redirect-feature/

Junnarkar, A., Adhikari, S., Fagania, J., Chimurkar, P., & Karia, D. (2021). E-Mail Spam Classification via Machine Learning and Natural Language Processing. *Proceedings of the Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks.*

Karim, A., Azam, S., Shanmugam, B., Kannoorpatti, K., & Alazab, M. (2019, Nov 20). A Comprehensive Survey for Intelligent Spam Email Detection. *IEEE Access*, pp. 168261 - 168295.

Kee, L. (2021, September 24). *Email And The Evolution Of Technology Standards*. Retrieved from Forbes.com: https://www.forbes.com/sites/forbestechcouncil/2021/09/24/email-and-theevolution-oftechnologystandards/?sh=1b86e6cefb25

Kenton, W. (2019, May 10). *Social Engineering*. Retrieved from https://www.investopedia.com/: https://www.investopedia.com//terms/s/social-engineering.asp

Kessel, P. v., & Allan, K. (2015). *Creating trust in the digital world EY's Global Information Security Survey 2015.* Ernst & Young Global Limited.

Khurana, D., Koli, A., Khatter, K., & Singh, S. (2017, Aug). Natural Language Processing: State of The Art, Current Trends and Challenges.

Knowles, C. (2020, Aug 20). *Australian cybersecurity company launches in NZ, plans for significant investment*. Retrieved from https://itbrief.com.au/: https://itbrief.com.au/story/australian-cybersecurity-company-launches-in-nz-plans-for-significant-investment

Kozierok, C. (2005, September 20). *IMAP Overview, History, Versions and Standards*. Retrieved from http://www.tcpipguide.com/: http://www.tcpipguide.com/free/t_IMAPOverviewHistoryVersionsandStandards-3.htm#:~:text=IMAP4%20is%20the%20current%20version,most%20recently%20by%20RFC%203501.

Lord , N. (2020, Dec 1). *Social Engineering Attacks: Common Techniques & How to Prevent an Attack*. Retrieved from https://digitalguardian.com/: https://digitalguardian.com/blog/social-engineering-attacks-common-techniques-how-prevent-attack

Mak, P. (2017, Feb 21). *The Human Factors in Cyber Security and Preventing Errors*. Retrieved from https://www.vircom.com/: https://www.vircom.com/blog/human-factors-in-cyber-security-preventing-errors/

Malware. (2021). Retrieved from https://www.cyber.gov.au/: https://www.cyber.gov.au/acsc/view-all-content/threats/malware

Mathews, L. (2019, Sep 6). *Toyota Parts Supplier Hit By $37 Million Email Scam*. Retrieved from https://www.forbes.com/: https://www.forbes.com/sites/leemathews/2019/09/06/toyota-parts-supplier-hit-by-37-million-email-scam/?sh=1bb6826f5856

Mohammed, S., Mohammed, O., Fiaidhi, J., Fong, S., & Kim, T. h. (2013). *International Journal of Hybrid Information Technology*, 43-56.

Musthaler, L. (2016, Mar 11). *How to use deep learning AI to detect and prevent malware and APTs in real-time* . Retrieved from https://www.networkworld.com/: https://www.networkworld.com/article/3043202/how-to-use-deep-learning-ai-to-detect-and-prevent-malware-and-apts-in-real-time.html

Nicholas, J. (2022, Octomber 1). *Types of Social Engineering Attacks: Detecting the Latest Scams*. Retrieved from https://www.biocatch.com/: https://www.biocatch.com/blog/types-social-engineering-attacks

O'driscoll, A. (2020, Oct 19). *What spear phishing is (with examples) and how you can avoid it*. Retrieved from https://www.comparitech.com/: https://www.comparitech.com/blog/information-security/spear-phishing/

Owaida, A. (2021, Jan 19). *FBI warns of voice phishing attacks stealing corporate credentials*. Retrieved from https://www.welivesecurity.com/: https://www.welivesecurity.com/2021/01/19/fbi-warns-voice-phishing-attacks-stealing-corporate-credentials/

Palmer, D. (2019, May 17). *What is GDPR? Everything you need to know about the new general data protection regulations*. Retrieved from https://www.zdnet.com/: https://www.zdnet.com/article/gdpr-an-executive-guide-to-what-you-need-to-know/

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., & Thirion, B. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research 12 (*, 2825-2830.

Petters, J. (2020, Mar 29). *What is a Whaling Attack?* Retrieved from https://www.varonis.com/: https://www.varonis.com/blog/whaling-attack/

Rose, A. (2021, January 21). *Most CISOs believe that human error is the biggest risk for their organization*. Retrieved from https://www.helpnetsecurity.com/: https://www.helpnetsecurity.com/2021/01/21/cisos-human-error-risk/

Rosenthal , J., & Oberly, D. (2020). FBI Warns Companies to Be Vigilant as COVID-19-Themed BEC Scams Continue to Grow. *The Computer & Internet Lawyer*, 6-8.

Sandler, R. (2020, Feb 27). *Shark Tank Host Barbara Corcoran Loses $380,000 In Email Scam*. Retrieved from https://www.forbes.com/: https://www.forbes.com/sites/rachelsandler/2020/02/27/shark-tank-host-barbara-corcoran-loses-380000-in-email-scam/?sh=59920c58511a

Satter, R., Donn, J., & Day, C. (2017, Nov 5). *Inside story: How Russians hacked the Democrats' emails*. Retrieved from https://apnews.com/: https://apnews.com/article/hillary-clinton-phishing-moscow-russia-only-on-ap-dea73efc01594839957c3c9a6c962b8a

Scammers taking advantage of COVID-19 to target small businesses. (2021, May 26). Retrieved from https://asic.gov.au/: https://asic.gov.au/about-asic/news-centre/news-items/scammers-taking-advantage-of-covid-19-to-target-small-businesses/

Schenkman, L. (2020, Jan 30). *Why we fall for phishing emails — and how we can protect ourselves*. Retrieved from https://ideas.ted.com/why-we-fall-for-phishing-emails-and-how-we-can-protect-ourselves/: https://ideas.ted.com/why-we-fall-for-phishing-emails-and-how-we-can-protect-ourselves/

(2015). *Science of Security (SoS) Initiative Annual Report.* The National Security Agency.

Security 101: Business Email Compromise (BEC) Schemes. (2016, Jan 11). Retrieved from https://www.trendmicro.com/: https://www.trendmicro.com/vinfo/us/security/news/cybercrime-and-digital-threats/business-email-compromise-bec-schemes

Shekhar, S. (2021, June 14). *www.analyticsvidhya.com*. Retrieved from LSTM for Text Classification in Python: https://www.analyticsvidhya.com/blog/2021/06/lstm-for-text-classification/

Shrestha, S. (2020, Jul 01). *NLP: Preparing text for deep learning model using TensorFlow2*. Retrieved from https://towardsdatascience.com/: https://towardsdatascience.com/nlp-preparing-text-for-deep-learning-model-using-tensorflow2-461428138657

Spear phishing. (n.d.). (trend micro) Retrieved 02 20, 2021, from https://www.trendmicro.com/vinfo/us/security/definition/spear-phishing

Stine, K., & Scholl, M. (2010). E-mail Security: An Overview of Threats and Safeguards. *Journal of AHIMA*, 28-30.

Stolfo, S., Hershkop, S., Wang, K., Nimeskern, O., & Hu, C.-W. (2003). A Behavior-Based Approach to Securing Email Systems. *Lecture Notes in Computer Science* (pp. 57-81). Berlin, Heidelberg: Springer.

Tandale, K., & Pawar, S. (2020). Different Types of Phishing Attacks and Detection Techniques: A Review. *2020 International Conference on Smart Innovations in Design* (pp. 295-299). Aurangabad, India: IEEE.

Taylor , T. (2019, May 13). *Why Email Is Still The Leading Communication Tool For Businesses*. Retrieved from https://techgenix.com/: https://techgenix.com/email-communication/

The 5 Most Common Types of Malware. (2020). Retrieved from https://www.checkpoint.com/: https://www.checkpoint.com/cyber-hub/threat-prevention/what-is-malware/the-5-most-common-types-of-malware/

(2014). *The digital economy, new business models and key features.* Paris: OECD Publishing. Retrieved from https://doi.org/10.1787/9789264218789-7-en

(2021). *The Global Risks Report 2021 16th Edition.* Geneva Switzerland: World Economic Forum. Retrieved from http://www3.weforum.org/docs/WEF_The_Global_Risks_Report_2021.pdf

The Phish Scale: NIST-Developed Method Helps IT Staff See Why Users Click on Fraudulent Emails. (2020, September 17). Retrieved from https://www.nist.gov/: https://www.nist.gov/news-events/news/2020/09/phish-scale-nist-developed-method-helps-it-staff-see-why-users-click

Tollefson, R. (2020, Nov 2). *BEC attacks: A business risk your insurance company is unlikely to cover*. Retrieved from https://resources.infosecinstitute.com/: https://resources.infosecinstitute.com/topic/bec-business-email-compromise-attacks-risk-insurance-company-wont-cover/

Types of Email Attacks and The Damage They Can Cause. (2016, Dec 15). Retrieved from https://www.cloudsecuretech.com/: https://www.cloudsecuretech.com/types-of-email-attacks-and-the-damage-they-can-cause/#:~:text=There%20are%20different%20types%20of,hackers%20to%20target%20email%20systems.&text=The%20most%20common%20techniques%20used,techniques%20that%20threaten%20email%2

Vaibhav, K., & Chandan. (2020, Sep 16). *Simple Mail Transfer Protocol (SMTP)*. Retrieved from https://www.geeksforgeeks.org/: https://www.geeksforgeeks.org/simple-mail-transfer-protocol-smtp/

Vardi, N. (2016, Feb 8). *How A Tech Billionaire's Company Misplaced $46.7 Million And Didn't Know It*. Retrieved from https://www.forbes.com/: https://www.forbes.com/sites/nathanvardi/2016/02/08/how-a-tech-billionaires-company-misplaced-46-7-million-and-didnt-know-it/?sh=2fb41d8d50b3

Verma, R., & Hossain, N. (2014). Semantic Feature Selection for Text with Application to Phishing Email Detection. *Information Security and Cryptology*, 455–468.

Volkman , E. (2019, Aug 8). *BEC Attacks: A Closer Look at Invoice Scams*. Retrieved from https://info.phishlabs.com/: https://info.phishlabs.com/blog/bec-attacks-invoice-scams

What is an Email Header? (2020). Retrieved from https://whatismyipaddress.com/: https://whatismyipaddress.com/email-header#:~:text=An%20email%20consists%20of%20three,which%20an%20email%20is%20routed.

Yaseen, K. Y., Abbas, K., & Ahmed, M. (2020). Image Spam Detection Using Machine Learning and Natural Language Processing. *JOURNAL OF SOUTHWEST JIAOTONG UNIVERSITY*.

Zorz, Z. (2019, Nov 19). *Your supplier's BEC problem is your BEC problem*. Retrieved from https://www.helpnetsecurity.com/: https://www.helpnetsecurity.com/2019/11/12/bec-problem/

Zweighaft, D. (2017). Business email compromise and executive impersonation: are financial institutions exposed? *JOURNAL OF INVESTMENT COMPLIANCE*, 1 - 7.

Appendix A: NLP Model Code

```
####Import Libraries####
import pandas as pand_l
import numpy as nump_l
import matplotlib.pyplot as plot_l
import re as reg_e
import string as str_v
import seaborn as sns
import matplotlib.pyplot as plt
import string
import tensorflow as ten_f

from keras_preprocessing.text import Tokenizer as token_processor
from keras_preprocessing.sequence import pad_sequences as pad_seq
from sklearn.feature_extraction._stop_words import ENGLISH_STOP_WORDS
from sklearn.model_selection import train_test_split as test_split
from sklearn.preprocessing import LabelEncoder as label_encounter

from sklearn.ensemble import RandomForestClassifier, GradientBoostingClassifier
from sklearn.metrics import precision_recall_fscore_support as score
from sklearn.metrics import accuracy_score as acs
from sklearn.metrics import confusion_matrix, classification_report
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split

######## Constant variables##########
CONST_SIZE_OF_VOCAB = 30000
CONST_EMBEDDING_DIMENSION = 32
CONST_EPOCHS = 15
CONST_MAX_LEN = 32
CONST_TRAINING_SIZE = 1000

CONST_TRUNC_TYPE = 'pre'
CONST_PADD_TYPE = 'pre'
CONST_UNK_TOK = '<UNK>'

##### Text Sanatization function #######

def remove_hyperlink(word):
    return  reg_e.sub(r"http\S+", "", word)

def to_lower(word):
    result = word.lower()
    return result

def remove_number(word):
    result = reg_e.sub(r'\d+', '', word)
    return result

def remove_punctuation(word):
    result = word.translate(str.maketrans(dict.fromkeys(str_v.punctuation)))
    return result

def remove_whitespace(word):
    result = word.strip()
    return result
```

```
def replace_newline(word):
    return word.replace('\n','')


def count_punct(text):
    count = sum([1 for char in text if char in string.punctuation])
    if len(text) > 0:
        return round(count/(len(text) - text.count(" ")), 3)*100
    return 0

def clean_up_pipeline(sentence):
    cleaning_utils = [remove_hyperlink,
                      replace_newline,
                      to_lower,
                      remove_number,
                      remove_punctuation,remove_whitespace]
    for o in cleaning_utils:
        sentence = o(sentence)
    return sentence

######### Read the corpa ##########

data = pand_l.read_csv('collectedcorpaspam.csv')
data["text"] = data.v2
data["spam"] = data.v1

v_emails_train, v_emails_test, v_target_train, v_target_test =
test_split(data.text,data.spam,test_size = 0.33)

data["body_text_clean"] = data.text.apply(lambda t: clean_up_pipeline (t) )
x_train = [clean_up_pipeline(o) for o in v_emails_train]
x_test = [clean_up_pipeline(o) for o in v_emails_test]

l_encounter = label_encounter()
y_train = l_encounter.fit_transform(v_target_train.values)
y_test = l_encounter.transform(v_target_test.values)

####Tokenization process#####

tokenizer = token_processor(num_words=CONST_SIZE_OF_VOCAB, oov_token=CONST_UNK_TOK)
tokenizer.fit_on_texts(x_train)

word_index = tokenizer.word_index


Training_Sequences = tokenizer.texts_to_sequences(x_train)
Training_pad = pad_seq(Training_Sequences, maxlen=CONST_MAX_LEN,
padding=CONST_PADD_TYPE, truncating=CONST_TRUNC_TYPE)

Testing_Sequences = tokenizer.texts_to_sequences(x_test)
Testing_pad = pad_seq(Testing_Sequences, maxlen=CONST_MAX_LEN,
padding=CONST_PADD_TYPE, truncating=CONST_TRUNC_TYPE)

#####Creating the model############

model = ten_f.keras.Sequential()
model.add(ten_f.keras.layers.Embedding(CONST_SIZE_OF_VOCAB,
CONST_EMBEDDING_DIMENSION, input_length=CONST_MAX_LEN))
model.add(ten_f.keras.layers.Bidirectional(ten_f.keras.layers.LSTM(64,
return_sequences=True)))
model.add(ten_f.keras.layers.Bidirectional(ten_f.keras.layers.LSTM(64)))
```

```
model.add(ten_f.keras.layers.Flatten())
model.add(ten_f.keras.layers.Dense(24, activation='relu'))
model.add(ten_f.keras.layers.Dense(1, activation='sigmoid'))

print(model.summary())

######Training the model ##################

Training_Sequences_padded = nump_l.asarray(Training_pad)
Testing_Sequences_padded = nump_l.asarray(Testing_pad)
Training_Labels = nump_l.asarray(y_train)
Testing_Lables = nump_l.asarray(y_test)

model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])
history = model.fit(Training_Sequences_padded, Training_Labels,
validation_data=(Testing_Sequences_padded, Testing_Lables), epochs=CONST_EPOCHS)

######Display model accuracy ##################
plot_l.plot(history.history['accuracy'])
plot_l.plot(history.history['val_accuracy'])
plot_l.title('model accuracy')
plot_l.ylabel('accuracy')
plot_l.xlabel('epoch')
plot_l.legend(['train', 'test'], loc='upper left')
plot_l.grid()
plot_l.show()

######Display model loss ##################
plot_l.plot(history.history['loss'])
plot_l.plot(history.history['val_loss'])
plot_l.title('model loss')
plot_l.ylabel('loss')
plot_l.xlabel('epoch')
plot_l.legend(['loss', 'test'], loc='upper left')
plot_l.grid()
plot_l.show()

######Creating confusion matrix about the model ##################

data['body_len'] = data['body_text_clean'].apply(lambda x: len(x) - x.count(" "))
data['punct%'] = data['body_text_clean'].apply(lambda x: count_punct(x))

X=data[['body_text_clean', 'body_len', 'punct%']]
y=data['spam']

X_train, X_test, y_train, y_test = train_test_split(X,y, test_size=0.2,
random_state=42)

tfidf_vect = TfidfVectorizer()
tfidf_vect_fit = tfidf_vect.fit(X_train['body_text_clean'])

tfidf_train = tfidf_vect_fit.transform(X_train['body_text_clean'])
tfidf_test = tfidf_vect_fit.transform(X_test['body_text_clean'])

X_train_vect = pand_l.concat([X_train[['body_len',
'punct%']].reset_index(drop=True),
           pand_l.DataFrame(tfidf_train.toarray())], axis=1)
X_test_vect = pand_l.concat([X_test[['body_len', 'punct%']].reset_index(drop=True),
           pand_l.DataFrame(tfidf_test.toarray())], axis=1)

X_train_vect.head()
```

```
rf = RandomForestClassifier(n_estimators=150, max_depth=None, n_jobs=-1)
rf_model = rf.fit(X_train_vect, y_train)
y_pred = rf_model.predict(X_test_vect)

precision, recall, fscore, train_support = score(y_test, y_pred, pos_label='spam',
average='binary')
print('Precision: {} / Recall: {} / F1-Score: {} / Accuracy: {}%'.format(
    round(precision, 3), round(recall, 3), round(fscore,3),
round(acs(y_test,y_pred)*100, 3)))

sns.set(rc= {"figure.figsize": (8, 6)})

cm = confusion_matrix(y_test, y_pred)
class_label = ["ham", "spam"]
df_cm = pand_l.DataFrame(cm, index=class_label,columns=class_label)
ax = sns.heatmap(df_cm, annot=True, fmt='d',cmap='YlGnBu')
bottom, top = ax.get_ylim()
ax.set_ylim(bottom + 0.5, top - 0.5)
plt.title("Confusion Matrix")
plt.xlabel("Predicted Label")
plt.ylabel("True Label")
plt.show()

######Model accuracy report ##################
report = classification_report(y_test, y_pred)
print(report)

##########Save the model############
model.save('pgb_spam_model')

######Test the model with real life spam text ##################
predict_text = 'take action today  up to  off on premium subscriptions give it a go
its time to fall in love you deserve it'
Test = tokenizer.texts_to_sequences([predict_text])[0]
Test_padded = pad_seq([Test], maxlen=CONST_MAX_LEN, padding=CONST_PADD_TYPE,
truncating=CONST_TRUNC_TYPE)
Test_padded = nump_l.asarray(Test_padded)
print('Done')
print(Test_padded.shape)

ypred = model.predict(Test_padded)

print (predict_text)
print("Spam percentage : {:.2f}%".format(100 * float(ypred[0])))
```