Monitoring Suspicious Discussions on Online Forum Using Deep Learning Techniques

Oluranti John AFOLABI
LCU/PG/000141

Being a MSc Thesis Submitted to the Department of Computer Science, Faculty of Natural and Applied Science, Lead City University, Ibadan, Oyo State, Nigeria

In Partial Fulfillment of the Requirements for the Award of Master of Science Degree (MSc) in Computer Science

2022

Certification

This is to certify that this thesis was carried out by Oluranti John AFOLABI with Matriculation Number LCU/PG/000141**,** a student in the Department of Computer Science under my supervision in the Faculty of Natural and Applied Science, Lead City University, Ibadan, Nigeria and that this has not been previously submitted.

…… ……

Azeez Ajani WAHEED (PhD) Date

Supervisor

…….... …….

Wilson SAKPERE (PhD) Date

Head of Department

Dedication

This project work is dedicated to God Almighty, the author and finisher of my faith who in His infinity mercy has kept me thus far.

Acknowledgement

My appreciation goes to the management and staff of Lead City University, for their support, aligns academic facilities and ambient environment provided during the course of study and in carrying out this study. With sincere heart I want to appreciate the wonderful support of my supervisor, Dr. Azeez Ajani Waheed who took pains in reading meticulously and offer constructive criticisms that help transform this report and the Head of Department, Dr. Wilson Sakpere has made a tireless effort to ensure the credibility of our research.

Indeed, I thank the Dean of Postgraduate School, Prof. Afolakemi Oredein for her guidance and support and as well as other staff of Postgraduate School. I want to express my gratitude to Dr. Phillip Achimugu who knows the beginning of it all and your staunch support with constant motivation really helped me a lot.

I am also greatly indebted to my loving and caring wife, Mrs. Toyosi Afolabi and my wonderful children: Seunfunmi Afolabi, Nifemi Afolabi and Feyikemi Afolabi for their kind support and understanding throughout the course of this study.

My gratitude also goes to all friends and colleagues who in one way or another has made this journey very memorable.

Finally, I appreciate the unquantifiable intellectual of several authors, scholars, and researchers whose works I've used. Although all the aforesaid people and institutions have helped in one way or other, all errors in this research report are solely mine if found.

Abstract

Nowadays, people are passionate about using the internet in their daily lives. This is rapidly increasing the use of online forums. An online forum is nothing but a medium to share one's thoughts, feelings and emotions towards specific pictures, videos and paintings etc. This leads to the execution of many legal and illegal activities. These illegal activities are trading black money online, distributing copyrighted movies and using illegal words. Law enforcement needs a system to effectively deal with this problem. Network Intrusion Detection System (NIDS) help system administrators detect network security vulnerabilities in their organization. However, many challenges arise when developing a flexible and effective NIDS for unplanned and unpredictable attacks. In this thesis, we propose a deep learning-based approach to develop a flexible and efficient NIDS to analyse suspicious and criminal activities occurring in forums. Deep learning technique is used, Natural Language Processing (NLP) for suspicious keyword extraction and Support Vector Machine (SVM) for detection and classification of suspicious keywords. We present the performance of our approach and compare it with some previous works. The metrics to be compared include accuracy, precision, recall, and f-measure values.

Keywords: online forum; machine learning; deep learning; natural language processing; support vector machine.

Word Count: 187

Table of Contents

Chapter Two: Literature Review

Chapter Four: Testing and Evaluation

Chapter Five: Conclusion

List of Tables

List of Figures

List of Appendices

List of Acronyms

Abbreviation		Meaning
AI	-	Artificial Intelligence
ANN	-	Artificial Neural Network
BBCode	-	Bulletin Board code
CSV	-	comma-separated values
DoS	-	Denial of Service
ER	-	Entity Relationship
FDI	-	Foreign Direct Investment
GB	-	Gigabyte
GII	-	Global Information Infrastructure
GPL	-	General Public License
HDD	-	Hard Disk
HTML	-	Hypertext Markup Language
ICT	-	Information and Communication Technology
IDS	-	Intrusion Detection System
IP	-	Internet Provider
kNN	-	k-Nearest Neighbors
ML	-	Machine Learning
MySQL	-	My Structured Query Language
NCWG	-	Nigeria Cybercrime Working Group
NLP	-	Natural Language Processing
OP	-	Original Post

PHP	-	Hypertext Preprocessor
PM	-	Private Message
R2L	-	Remote-to-local
RAM	-	Read-Access Memory
RDBMS	-	Relational Database Management System
SIEM	-	Security Information and Event Management
SQL	-	Structured Query Language
SVM	-	Support Vector Machine
TB	-	Terabyte
TCP	-	Transmission Control Protocol
U2R	-	UsertoRoot
UDP	-	User Datagram Protocol

Chapter One

Introduction

1.1 Background to the Study

Internet has become an effective and convenient communication channel for knowledge shearing, opinions expression, products advertisement, and post textual data via the browser interface to share information with each other[1]. It is important to extract useful information from this plain text data to reveal the hidden data. The goal of machine learning is to extract information from large datasets and transform it into an easy-to-understand format. As Internet technology continues to grow, it has led to lawful and unlawful activities. Many direct messages were discussed in Internet forums long before they were published in the traditional media. This communication channel provides an effective channel for unlawful activities such as copyrighted movie distribution, threatening messages, and online gambling.

Hackers often use social media networks to discuss cyber-attacks, share strategies and tools, and identify potential victims for targeted attacks. Analysts examining these discussions can forward information about malicious activity to system administrators who can then detect, defend against, and recover from future attacks. For example, prior to the anticipated cyber-attacks on Israeli government websites by the hacking group Anonymous, government analysts were monitoring hackers on Facebook and in private chat rooms. As a result, system administrators were prepared to counter distributed denial-of-service attacks and defacement of government websites[2]. In addition, many facts prove that simply managing information on the Internet through traditional management models is not enough. In this regard, web mining

is a new research direction for information gathering and analysis on the explosive and unstructured Internet. Criminal web data always provides valuable and relevant information for legal management purposes. It is always very difficult to assess the various capabilities of a wide range of criminal web data and is therefore one of the most notable tasks of legal management.

Crime can be as extreme as murder or rape and requires sophisticated analytical techniques to extract useful information from the data[3]. Crime is a serious problem in today's world. One of the factors influencing the number of crimes is the technological advances we face globally. Crimes can include theft, Internet crime, theft and fraud. In general, people commit crimes out of greed, anger, jealousy, revenge or pride, which affects other law-abiding citizens[4]. In addition, when a crime is committed, the organizational structure and management can be disrupted, and security measures can be provided in all areas of our lives. This criminal activity can be put to an end through suspicious chat detection system[5].

In recent decades, information technology has made great strides in hardware and software. This has had a huge impact on simplifying the communication processes of organizations, especially in the business arena. Originating in the United States, the Internet is based on "a network of connected computers, each connected to a set of other computers that communicate electronically between computers around the world" (Henslowe, 1999, p.87). To better understand the evolution of the internet and the advent of social media[6]. A social network is an interactive platform where individuals on the network create, share and share content[7].

Forums are a platform for people from different geographic regions to express and share their opinions, and through marketing and communication they influence many aspects of their lives. Monitoring suspicious discussions on forums is the best way to measure user loyalty. Some people use these discussion forums for illegal purposes by posting suspicious chats in the form of text, video or images and sharing them with other users[8].

Chat can be understood as the process of communication, interaction and/or messaging over the Internet. This includes two or more people communicating through chat support services. Chat can be done via text and voice communication. This may include live or online chat[9].

Machine learning is a field of Artificial Intelligence (AI) and computer science, with a focus on using data and algorithms to imitate human learning methods and gradually improve accuracy[10]. The goal of machine learning is to extract information from large datasets and transform it into an easy-to-understand format.

Deep Learning is a type of machine learning that trains computers to perform human-like tasks such as: Speech recognition, image identification, or prediction. Instead of organizing the data to execute predefined equations, deep learning sets basic parameters for the data and uses many layers of processing to recognize patterns so that the computer can recognize them. Train to learn on your own[11].

1.2 Statement of the Problem

This data actually means a huge virtual space that anyone can discuss in the form of posted messages. User preferences are generally captured by analyzing their attitudes and behaviors mentioned on the websites as comments. To measure a user's loyalty and to keep track of their sentiments towards any topic is achieved by monitoring the suspicious activities and discussions done through their posts on the social media[12].

Online chat can imply any type of communication over the Internet that provides continuous sending of instant messages from the sender to the recipient[13]. It is vulnerable to security attacks and there is no encryption standard for chat conversations and the system was unable to handle big data[14]. The system was limited to six keywords for suspicious detection[15]. The main hurdle faced by researchers is lack of information retrieval and data analysis tools for real time data[16]. the recognition of the suspicious words using automated surveillance is a big challenge[17]. The challenging is because of the exponential development in information and communication technology[8].

Several works have been done addressing monitoring suspicious discussions on online forum, but still need to develop an improved system due to the limitations of the existing systems which are:

i. No restriction for sending suspicious messages
ii. Malicious users will be blocked only after the account has been reported.
iii. Admin permission is needed before any user can be added to the forum.

1.3 Aim and Objectives

The aim of this research is to design a system that monitors suspicious discussions using deep learning technique.

The objectives are as follows:

i. to design a model with a chatbox that will enable an admin to monitor suspicious discussions online.
ii. to implement the model designed in (i) above
iii. to evaluate the performance of the implemented system

1.4 Research Questions

The research questions of the study are derived from the objectives of the study as stated below:

1. Why are the existing technologies of Chat Monitoring Systems still inefficient and unable to protect chat forum users from being threatened?
2. How should chat forum users be prevented from being threatened?
3. How should monitoring suspicious discussion system be implemented to prevent perpetrators.
4. How should monitoring suspicious detection system be designed and implemented?
5. How should a deep learning techniques be used to improve suspicious detection accuracy?

As stated above, this research work includes understanding and knowing the limitation of Monitoring Suspicious Discussion on Online Forum.

1.5 Significance of the Study

This research will help to detect any suspicious posts or comments on online forum, classify and monitor any suspicious text comments online and to stop threatening messages in social Media.

1.6 Scope of the Study

A chat forum is developed using Python programming language, Natural Language Processing (NLP) is used for suspicious keyword extraction from Online Forum Chat dataset and Support Vector Machine (SVM) is used for detection and classification of suspicious keywords.

1.7 Limitations of the Study

This study will employ the use of just one of the machine learning techniques (Deep Learning) to monitor suspicious discussions online, as there are lots of machine learning techniques. Future works can be done by implementing other machine learning technique and then have a comparison of which of these technique works best for monitoring suspicious discussions online.

1.8 Operational Definition of Terms

Online Forum: A facility on a computer network (especially the Internet) that allows users to participate in discussions and exchange information and opinions on specific topics, especially websites dedicated to such discussions.

Artificial Intelligence (AI): The theory and development of computer systems capable of performing tasks that normally require human intelligence, such as vision, speech recognition, decision making, and translation between languages.

Artificial Neural Network (ANN): This is a computer architecture modeled after the human brain, consisting of nodes connected by links of varying strength.

Crime: This is the intentional conduct of an act that is generally considered socially harmful or dangerous and is clearly defined, prohibited, and punishable by criminal law.

Internet: This is a global computer network consisting of networks that provide a variety of information and communication functions and are interconnected using standardized communication protocols.

Intrusion Detection System (IDS): Is a surveillance system that detects suspicious activity and generates alerts upon detection. Based on these alerts, Security Operations Center (SOC) analysts or incident responders can investigate the issue and take appropriate action to remediate the threat.

k-Nearest Neighbors (kNN): This is also known as ANN or k-NN, is a nonparametric supervised learning classifier that uses proximity to make classifications or predictions about the clustering of single data points.

My Structured Query Language (MySQL): MySQL HeatWave is a fully managed service that enables customers to run OLTP, OLAP, and machine learning workloads directly from their MySQL Database.

Hypertext Preprocessor (PHP): It is a server-side scripting language that can be embedded in HTML documents and is used to run applications on servers to deliver content and

applications over networks or the Internet. The PHP scripting engine can be installed and run on a single computer or server.

Support Vector Machine (SVM): Support vectors are data points near a hyperplane that affect the position and orientation of the hyperplane. Use these support vectors to maximize the classifier margin. Deleting support vectors changes the position of the hyperplane. These are the points that help build the SVM.

Natural Language Processing (NLP) refers to the branch of computer science—and more specifically, the branch of artificial intelligence or AI—concerned with giving computers the ability to understand text and spoken words in much the same way human beings can.

Transmission Control Protocol (TCP): Is a communication standard that allows application programs and computing devices to exchange messages over a network. It is designed to send packets over the Internet and reliably deliver data and messages over networks.

Endnotes

1. K. Ishii, M.M. Lyons, & S.A Carr, Revisiting Media Richness Theory for Today and Future. *Human Behavior and Emerging Technologies,* 1(2), 2019, pp.124-131.

2. C.J. Howell, G.W. Burruss, D. Maimon & S. Sahani, Website Defacement and Routine Activities: Considering the Importance of Hackers' Valuations of Potential Targets. *Journal of Crime and Justice,* 42(5), 2019, pp.536-550.

3. H. Tariq, M.K. Hanif, M.U. Sarwar, S. Bari, M.S. Sarfraz & R.J. Oskouei, Employing Deep Learning and Time Series Analysis to Tackle the Accuracy and Robustness of the Forecasting Problem. *Security and Communication Networks,* 2021.

4. P. Paul, & P.S. Aithal, Cybercrime: Challenges, Issues, Recommendation and Suggestion in Indian Context. *International Journal of Advanced Trends in Engineering and Technology. (IJATET),* 3(1), 2018, pp.59-62.

5. A. Shah & D. Chudasama, Investigating Various Approaches and Ways to Detect Cybercrime. *Journal of Network Security,* 9(2), 2021, pp.12-20p.

6. V.A. Briciu & A. Briciu, Social Media and Organizational Communication. *Encyclopedia of Organizational Knowledge, Administration, and Technology,* 2021, (pp. 2609-2624). IGI Global.

7. N. Enke & N.S. Borchers, Social Media Influencers in Strategic Communication: A Conceptual Framework for Strategic Social Media Influencer Communication. *Social Media Influencers in Strategic Communication,* 2021, (pp. 7-23). Routledge.

8. J.K. Thomas, M. Redhya & T. Mahalekshmi. Monitoring Suspicious Discussions in Online Forums Using Data Mining. *International Journal of Research Publication and Reviews,* Vol 2, no 12, 2021, pp 682-688, ISSN 2582.

9. J. Chen & P. Neo, Texting the Waters: An Assessment of Focus Groups Conducted Via the Whatsapp Smartphone Messaging Application. *Methodological Innovations,* 12(3), 2019, p.2059799119884276.

10. K. Aggarwal, M.M. Mijwil, A.H. Al-Mistarehi, S. Alomari, M. Gök, A.M.Z. Alaabdin & S. H. Abdulrhman, Has the Future Started? The Current Growth of Artificial Intelligence, Machine Learning, and Deep Learning. *Iraqi Journal for Computer Science and Mathematics,* 3(1), 2022, pp.115-123.

11. Y.K. Wong, Advanced Deep Learning Approach and Applications. *International Journal of Information Technology (IJIT),* 7(5). 2021.

12. R. Rawat, V. Mahor, S. Chirgaiya & A.S. Rathore, Applications of Social Network Analysis to Managing the Investigation of Suspicious Activities in Social Media Platforms. *Advances in Cybersecurity Management,* 2021, (pp. 315-335). Springer, Cham.

13. S.B. Avaghade, A.V. Rajkumar, P.J. Jayvant, S.S. Prithviraj, P.A. Dilip & K. Dacoe, Active Chat Monitoring and Suspicious Detection Over Internet. *International Journal of Engineering Applied Sciences and Technology,* Vol. 4, Issue 11, ISSN No. 2455-2143, 2020, pages 397-399

14. B. Ahmad, W. Jian & Z. Anwar Ali, Role of Machine Learning and Data Mining in Internet Security: Standing State with Future Directions. *Journal of Computer Networks and Communications,* 2018.

15. T. Srivastava, R. Mangalagowri & S.S. Dudala. Monitoring of Suspicious Discussions on Online Forums Using Data Mining. *International Journal of Pure and Applied Mathematics,* 118(22), 2018, pp.257-262.

16. A. Kumari & Balkishan, Detection of Suspicious Text Messages and Profiles Using Ant Colony Decision Tree Approach. *International Journal of Business Intelligence and Data Mining,* 19(4), 2021, pp.418-442.

17. M. Brindha, V. Vishnupriya, S. Rohini, M. Udhayamoorthi & K.S. Mohan, *Active Chat Monitoring and Suspicious Detection Over Internet,* 2019.

18. J. Risch & R. Krestel, Toxic Comment Detection in Online Discussions. *Deep Learning-Based Approaches for Sentiment Analysis,* 2020, (pp. 85-109).

19. A. Habib, F. Hossain, T. Ferdous & K.M. Bayezid, Social Networks and Social Ties: Changing Trends Among Urban Dwellers in Bangladesh. *Open Access Library Journal,* 5(5), 2018, pp.1-12.

Chapter Two

Literature Review

The chapter provides an overview of existing research on monitoring suspicious discussions on online forum. It introduces the framework for the case study that comprises the main focus of the research described in this thesis. Before the literature review, it is necessary to first provide a brief explanation on online forum and its application, intrusion detection, machine learning techniques and deep learning.

2.1 Conceptual Review

2.1.1 Online Forums

An Internet forum or bulletin board is an online discussion site where people can talk in the form of posted messages. It differs from chat rooms in that messages are longer than lines of text and are at least temporarily stored. Depending on user access level and forum configuration, posted post may require approval from moderators before it can be made public[1]. Forums are associated with specific terminology. Example: A single conversation is called a "thread" or topic.

Discussion forums have a hierarchical or tree structure. Forums can contain multiple subcategories, and each subcategory can contain multiple topics. In a forum thread, each newly started discussion is called a thread and can be answered by any number of people. Depending on forum settings, users can be anonymous or register to the forum, then log in to post. Most forums do not require users to log in to read existing posts.

2.1.2 Online Forums History

The modern discussion board is derived from the bulletin board, and is also known as the laptop conference structure, and is a technological evolution of the dial-up bulletin board facility. From a technology perspective, forums or forums are neat packages that handle user-generated content[2].

Internet forums can initially be defined as an Internet version of a digital mailing list or discussion group (which exists on Usenet); allows people to display messages and access different messages. Later trends mimic newsgroups or one-on-one lists, which have multiple discussion forums devoted to a particular topic.

Internet signage is the norm in many developed countries. Japan publishes at most [citation needed] over millions per day on the country's largest discussion forum, 2channel. China also has several hundred thousand posts on forums, including the Tianya Club.

Some of the earliest discussion forum structures are the PlanetForum widget, developed in the early 1970s, the ESIA widget, which first became operational in 1976, and the KOM widget, which first became operational in 1977.

One of the primary discussion board sites (which nonetheless energetic today) is Delphi Forums, as soon as known as Delphi. The service, with 4 million members, dates to 1983. Forums carry out a feature much like that of dialup bulletin board structures and Usenet networks that have been first created beginning withinside the overdue 1970s. Early

netprimarily based totally boards date again as a long way as 1994, with the WIT mission from W3 Consortium and beginning from this time, many options have been created[3]. A feel of digital network regularly develops round boards which have everyday users. Technology, video games, sports, music, fashion, religion, and politics are famous regions for discussion board themes, however there are boards for a big quantity of topics. Internet slang and photo macros famous throughout the Internet are ample and broadly utilized in Internet boards.

Forum software program applications are broadly to be had at the Internet and are written in quite a few programming languages, together with PHP, Perl, Java and ASP. The configuration and information of posts may be saved in textual content documents or in a database. Each package deal gives oneofakind functions, from the maximum basic, presenting textual contentbest postings, to extra superior applications, presenting multimedia aid and formatting code (typically called BBCode). There are many applications that can be seamlessly incorporated into an existing website to allow traffic to post comments on articles.

Several different Internet packages, including blogging software, also include discussion forum functions. WordPress comments at the bottom of a blog post allow a single-threaded chat of any blog post. Slashcode, on the other hand, is much more complex, allows for complete streams, and incorporates a powerful moderation and moderation engine in addition to some of the profiling functionality available to forum users[4].

Several independent forum topics have gained fame and popularity, including the "I'm lonely, we'll all talk to me" thread on the MovieCodec.com forum, which has become a "retreat place" to the lonely[5].

2.1.3 Structure Online Forums

Forums are organized in a directory tree. The best conclusion is "genre". Forums can be divided into categories for meaningful dialogue. Below the categories are sub-forums, and these sub-forums help you create more sub-forums. Topics (commonly referred to as threads) fall into the lowest levels of sub-forums, where individuals can start discourses and posts. Forums are coherently organized on a limited number of non-specific topics (usually basic themes), directed and reviewed by a group of famous members and managed by a group called moderators[10]. It can also be a graphic structure. All bulletins use one of three possible display formats. Each of the three main message board display formats: unthreaded / half-threaded / fullthreaded has its own strengths and weaknesses. If the messages are completely irrelevant, then the wireless format is best. If the client has a message point and that message point has multiple responses, the semi-threaded format is best. If the client provides a message point and responds in response to the subject of that message, stream format is completely preferred at this stage [6].

2.1.4 User's Group

The Western style forum organizes guests and logged-in individuals into user groups. Benefits and rights are granted based on these groups. As a result, forum users can move to more favorable user groups based on the criteria set by the administrator. People who view

closed threads as part will see a box stating that they don't have the right message, but arbitrators may see the same box where they can post more than just news[7].

Unregistered User Locations commonly referred to as visitors or guests. Visitors have regular access to all features that do not require protection from database changes or security breaches. In most cases, visitors can view the content of the forum and use highlights such as skip stamps, but if the guest can log in, the administrator can refuse to review the forum. People who frequently access forums, sections, and even threads are called larkers, and the habit is called larking.

2.1.5 Administrator

The administrator (abbreviation: "admin") takes care of the points of special interest needed to manipulate the location. As a result, you can promote (and demote) people to and from moderators, monitor rules, create sections and subsections, and perform all database operations (such as database backups). Administrators act as moderators on a regular basis. Administrators can also create statements for the entire forum and change the appearance of the forum (called skins). There are too many forums for admins to share ideas[8].

2.1.6 Post

Posts are messages sent by the user and are enclosed in a square containing the points of interest to the user and the date and time they were sent[9]. Members can modify or delete their posts. Contributions are included in the string and are displayed one after the other as blocks. The first post starts a thread. This is sometimes referred to as TS (thread starter) or OP

(original post). Posts that follow the thread mean to continue the discourse about the post or respond to other replies. Conversations often fail.

In the Western Forum, the classic way of subtle expressions (such as member titles and avatars) was the left side of the post, the fixed width contract column, and the right side post control. It is located above the signature line at the bottom of the main unit. Recent use of the Forum Computer Program has duplicated the Asian way of displaying subtle elements of members above posts.

Posts have their own limits, which are usually measured in characters. You need a message that is at least 10 characters long on a regular basis. There is always an upper limit, but it is sometimes achieved-most sheets are either 10,000., 20,000., 30,000., or 50,000. characters[10]. Most forums track user posts. Posts is an estimated number of posts created by a particular user. Users with high posts are regularly considered more legitimate than users with low posts, but they are inconsistent. Few forums have disabled post recounting based on the belief that quality of information is more important than quantity.

2.1.7 Thread

A thread may be a collection of posts, usually displayed from oldest to most recent, in spite of the fact that this is often regularly configurable: Choices for recent to oldest and for a threaded view (a tree-like view applying logical reply structure before chronological order) can be available.[11] A thread is characterized by a title, an extra portrayal which will summarize the intended discussion, and an opening or unique post (common abbreviation OP,

which can also mean original poster), which opens anything discourse or makes anything declaration the poster wished. A thread can contain any number of posts, counting different posts by the same person if they are consecutive.

2.1.8 Stickying

Threads that are important but rarely receive posts are stickyed (or, in some software, "pinned"). A sticky thread will always appear in front of normal threads, often in its own section. A "threaded discussion group" is simply any group of individuals who use a forum for threaded, or asynchronous, discussion purposes. Groups may or may not be the only users of the forum.

2.1.9 Discussion

Forums prefer the Openo standard assumptions. The most common topics in the forum include questions, comparisons, votes, and discussions. It's not uncommon for irrational or antisocial behavior to emerge when people lose their cool, especially when the subject is controversial. Not understanding the difference in value between participants is a common problem in forums[12].

Responses to topics often focus on someone's point of view, so discussions often tend to be multi-directional when people question the other's validity or sources. Circular threads and ambiguity in response can span dozens of posts on the topic and eventually end when people give up or attention span fluctuates and a more interesting topic emerges. It is not uncommon for an argument to end in a personal attack.

2.1.10 Owners liabilities and moderators

Several defamation and compensation proceedings have been filed against forums and moderators. The recent proceeding is the Scubaboard proceeding in which a Maldivian company filed a lawsuit against Scubaboard in January 2010 for defamation and defamation[13].

2.1.11 Online Forums Common Features

To become an Internet forum, web applications need the ability to send threads and responses by default. Normally, the thread is in the new view and the response is from the old view to the new view.

2.1.12 Tripcodes and Capcodes

Most 2-channel photograph forums and dialogue boards allow (and encourage) nameless posting and use the tripcode device as opposed to registration.

The tripcode is the end result of a password hash that permits you to perceive your identification while not having to keep any records approximately the user. The ride code device provides a mystery password after the username delimiter (regularly more than a few sign)[14]. This password or ride code is hashed to a unique key or ride that may be prominent from the call the use of HTML style. Tripcodes can't be forged, however a few styles of discussion board software program are insecure and may be guessed. For different types, it is able to be brutally enforced the use of software program designed to locate ride codes, which include Tripcode Explorer.

Moderators and directors regularly assign every different a cap code or ride code wherein the inferred ride is changed with a unique note (which include "# Administrator") or cap.

2.1.13 Private Message (PM)

Private Messages (PMs) are private messages sent by one member to one or more other members. You can send so-called hidden copies. When sending a blind copy (bcc), the user to whom the message is sent directly does not know who the recipient of the blind copy is or if it has been sent[15].

Private messages are often used for private conversations. It can also be used with a travel code. The message is sent to a public route and can be retrieved by entering the route code.

2.1.14 Attachment

Attachments can be any file. When someone attaches a file to someone's post, that particular file is uploaded to the forum server. Forums often have very strict limits on what you can and cannot participate in (including the size of the file in question). Attachments can be part of a thread or a social group[16].

2.1.15 Machine Learning

Machine Learning (ML) is a scientific study of algorithms and statistical models that computer systems use to perform specific tasks by relying on patterns and inferences without the use of explicit instructions. It is considered a subset of artificial intelligence. Machine

learning algorithms build mathematical models based on sample data called "training data" and make predictions and decisions without explicitly programming them to perform tasks. Machine learning algorithms are used in a variety of applications, such as email filtering and computer vision when it is difficult or impossible to develop traditional algorithms to do the job effectively.[17]

Machine learning is closely related to computational statistics that focus on making predictions using computers. Mathematical optimization research provides methods, theories, and application domains in the field of machine learning. Machine learning is also known as predictive analytics.

2.1.16 Deep Learning Technique

Deep learning is one of the cornerstones of artificial intelligence (AI). It is used for image classification, speech recognition, object recognition, and content description. Deep learning technology has improved the ability to classify, recognize and describe. Deep learning is a type of machine learning that trains computers to perform human tasks such as: language recognition, image recognition, or prediction. Instead of organizing data to go through predefined equations, deep learning identifies the basic parameters of the data and uses multiple layers of processing to recognize patterns so that computers can recognize it.[18]

2.1.17 Intrusion Detection System

Intrusion Detection System is a device or software application that monitors your network or system for malicious activity or policy violations. Intrusion activities or breaches are

typically reported to the administrator or centrally collected using Security Information and Event Management (SIEM). The SIEM system combines outputs from multiple sources and uses alert filtering techniques to distinguish between malicious activity and false positives.

IDS type range in scope from unmarried computer systems to massive networks[19]. The maximum not unusual place classifications are community intrusion detection structures (NIDS) and host-primarily based totally intrusion detection structures (HIDS). A machine that video display units essential working machine documents is an instance of an HIDS, whilst a machine that analyzes incoming community site visitors is an instance of an NIDS. It is likewise feasible to categorise IDS via way of means of detection approach. The maximum famous editions are signature-primarily based totally detection (spotting horrific patterns, consisting of malware) and anomaly-primarily based totally detection (detecting deviations from a version of "good" site visitors, which regularly is predicated on system learning). Another not unusual place variation is recognition-primarily based totally detection (spotting the capacity risk consistent with the recognition scores). Some IDS merchandise have the potential to reply to detected intrusions. Systems with reaction skills are commonly called an intrusion prevention machine[20]. Intrusion detection structures also can serve particular functions via way of means of augmenting them with custom tools, consisting of the usage of a honeypot to draw and signify malicious site visitors.

2.1.18 Network Intrusion Detection Using Machine Learning Techniques

With the rapid development of techniques used to attack these networks, cybersecurity concerns have increased in recent years. The techniques used in these attacks attempt to use

network packets that have similar characteristics to normal traffic, making it increasingly difficult to detect traffic coming from intrusion attempts, unlike like traditional network protection techniques. It is very vulnerable to such attacks. As a result, more sophisticated techniques are being developed to protect these networks against complex attacks such as: using machine learning to distinguish normal traffic packets from attack packets[21].

Machine learning is a field of study that aims to enable computers to acquire knowledge about the outside world without human intervention. The knowledge extracted by a particular machine learning technique may differ from the outside world for a different set of inputs. Furthermore, the knowledge extracted from a set of inputs can differ from one machine learning method to another, depending on the different approaches used for extraction. export this knowledge. One of the main areas of machine learning is data mining and input from the outside world[22].

Data mining and other machine learning techniques fall into two main categories: unsupervised and unsupervised. Unsupervised data mining does not need to be added to the input dataset. The purpose of these techniques is to extract relationships between objects in a data set, but unsupervised data mining requires an expert to add additional information to the data set. Supervised data mining technique extracts relationships between objects in a data set and knowledge added by experts. This knowledge, extracted from a sample dataset known as the training dataset, can be used at runtime to apply the extracted knowledge to new objects and support operation of systems that interact with the domain[23].

Classification is one of the most widely used data mining techniques where information is added to a data set in the form of labels that categorize each object in the dataset into one or more classes that reside in the field. During the training phase of the classifier, the attributes of the objects of each class are extracted and the classifier builds a model based on these relationships. These models are then used to predict new classes of data objects at runtime to support system decision making in domain management. These decisions are based on the characteristics of the category for which the data object is expected. Therefore, different intrusion detection systems (IDS) have been proposed based on the classification method. The system trains the classifier to collect data from normal packets and network traffic, including attack packets, so you can predict the class of each object containing attack packets[24].

2.1.19 Network Traffic Datasets

To generate and evaluate the performance of data mining techniques in the field of network protection, various data sets are collected based on the packet characteristics of network traffic, including both normal and packetized packets. attack. Since classifiers have different approaches to extracting knowledge from a dataset, it is important to evaluate classifier performance as a measure of the quality of the extracted knowledge. However, since the classifier is used to provide predictions, labeled data is used for this evaluation by comparing the predictions provided by the classifier with the actual class or label that the classifier provides. These objects belong to. Therefore, each data set is divided into two parts, one part used by the classifier to extract knowledge is called the training dataset, and the other part is the prediction class with the actual class being evaluated[25].

One of the first datasets collected for network traffic to train and evaluate data mining techniques at IDS was the KDD Cup 99 dataset[26]. This dataset contains information about 4 898,431 network packets, each package is characterized by 41 different characteristics. Each package is labeled with one of five labels. One is for the regular packages and the other four are for the various cyberattacks included in the record, namely:

1. Tracking attack: An attack that collects all possible information about a network and the computers that belong to it and uses this information to compromise the security of the network.
2. UsertoRoot (U2R) Attack: An attack that exploits the information of users who have legitimate access to this network to gain root access to the system by exploiting a vulnerability in this system.
3. Denial of Service (DoS) attack: An attack that attempts to use up available resources on a computer, such as memory and processing power, to deny service to legitimate users.
4. Remote-to-local (R2L) attack: An attack in which an attacker can gain access to a network but does not have the necessary credentials to authenticate the services provided by this network.

This dataset is commonly used to train and evaluate many intrusion detection systems based on data mining and the various problems this dataset encounters under[27]. The first problem is mentioned in this log is the l tool used to collect packet information such as: unchecked when collecting datasets. Another issue is the definition of dataset attacks. For example, a probe attack is only considered a true attack if a certain threshold is exceeded if such conditions are

not taken into account in data collection. In addition, the number of redundant objects in the data set is very large and it is very likely that similar objects in the test dataset will be found in the training dataset, which will affect the parsing difficulty and these objects. Reduce class's hard to predict. Object[28].

The issue of data object redundancy in the dataset has been resolved in this release, but there are concerns about applying the aggregated data to real-world domains and environments used. used to collect this data. The configuration qualifies as suspicious[29].

An up-to-date dataset of network traffic, including regular and attacked packets, known as the UNSWNB15 dataset, has been proposed[30]. This dataset consists of 2,540,047 datasets, each representing a network packet described with 47 attributes. On two labels. One label represents the state of the packet (normal packet or hacked packet), and the other, for the attack packet, represents the attack pattern of the nine attacks containing the packet. These attacks are Dos, Exploit, Backdoor, Generic, Fuzzers, Analysis, Reconnaissance, Shell Code, Worms[31].

2.1.20 Machine Learning Techniques

Various machine learning techniques are used to implement the intruder detection system using the data set presented above. Firewalls are network components that analyze packet information to determine whether to allow or deny packets access to your network. So use these techniques with these firewalls to protect your network. The information in each packet is pulled from the firewall and sent to machine learning to predict whether the traffic is

normal or under attack. These predictions are used to make and enforce firewall decisions. This section describes machine learning techniques used in intrusion detection systems[32].

The kNN classifier is a lazy classifier that doesn't extract any knowledge from the training dataset until a prediction is requested from the classifier. This knowledge is extracted whenever a prediction is required by retrieving the k most similar data objects from the training dataset and choosing the dominant class of these objects. The dominance class can be calculated by voting, i.e. the class with the highest number of data objects, or by weighting each layer using the distance between the new feature on one side and each feature in the training data set from another object. Since the distance between two objects is affected by the number of strokes that characterize these objects, it is important to reduce the number of strokes to the minimum number possible, to reduce the time required to calculate the distance. way and speed up computation. To do this, each feature is ranked according to its role in the classification process and the lowest ranked features are discarded or optimization algorithms are used to remove features that have no effect or have a negative effect on classification results[33].

The decision tree classifier uses a set of IF/THEN clauses distributed across multiple levels in a tree distribution, depending on the results of a comparison made at a particular level and a comparison. comparison is made at the previous level. Let it be decided. These sets are generated during the training phase of the classifier and applied to the feature values of the new data object to predict each class. Features that have a significant impact on class prediction are placed above the decision tree and closer to the root of the tree. Decision tree

has its root at the top[34]. In addition, some techniques use multiple trees to provide better classification results. Each tree is trained in a different set of random training data. Therefore, such classifier is called random forest, and the prediction class is the dominant class among the predicted tree classes in the forest[35]. Figure 1 shows an example of a decision tree to predict a player's state of whether or not to play today, based on the weather forecast.

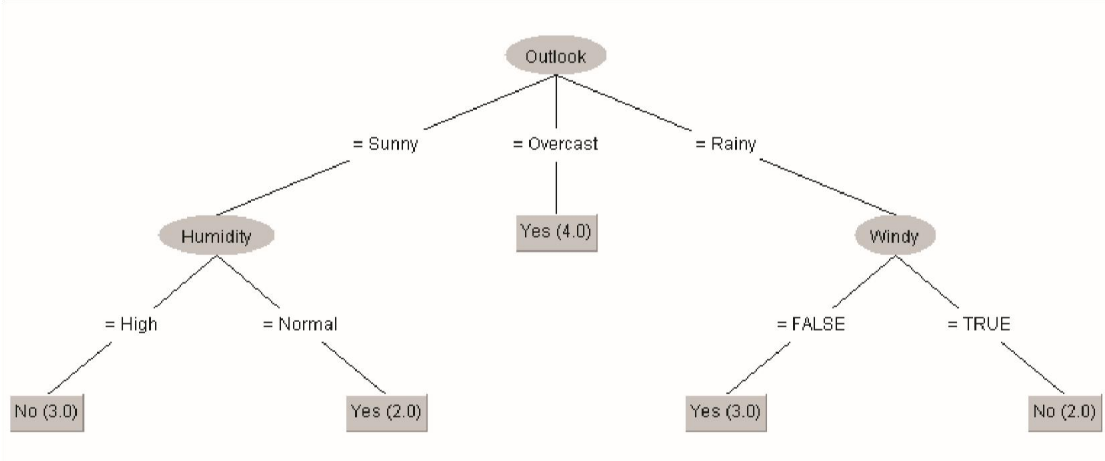

Figure 2.1. An example of a Decision Tree Based On a Weather Forecast.

Sources: Online (2021)

Support vector machines (SVMs) are another popular classifier used in intrusion detection systems. This classifier distributes the objects of the training dataset in an n-dimensional space. Where n is the number of objects in the data set. Boundaries between regions containing objects of the same class are optimized by the SVM classifier to optimize the distance between the boundaries and the nearest feature of each adjacent class. This makes the prediction more reliable because the reliability of the SVM prediction is described by the distance from the new data object to which the class is predicted and the limit of this class[36]. Figure 2 shows an example of an SVM constraint in a three-layer two-dimensional space.

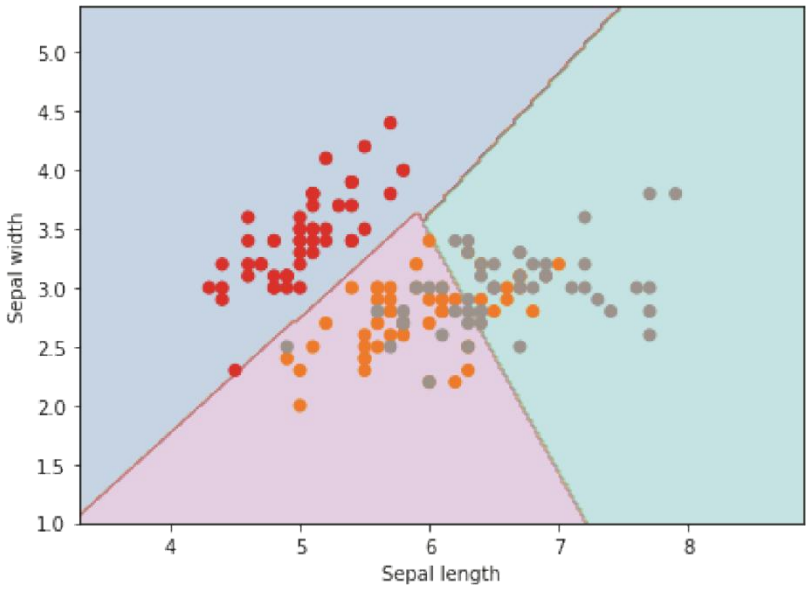

Figure 2.2. An Example of an SVM Boundary in 2D Space.

Sources: Online (2021)

Recently, artificial neural networks (ANNs) have received considerable interest among other machine learning techniques due to their relatively good performance, especially on large data sets. Artificial neural networks are mathematical representations of the connections between neurons in the human brain and the neurons that guide the decisions humans make. These neurons are distributed in layers, and there are three types of layers in the ANN. The first type of layer is the input layer; whose number of neurons is equal to the number of features that characterize the dataset. The second type of layer is the output layer; whose number of neurons is equal to the number of outputs provided by the network. The topology of these layers is controlled by how these networks connect the external domains, so a third type of layer, the hidden layer, is added to these networks, giving flexibility to the topology. overall ANN results. It will increase. The number of hidden layers and the number of neurons in each layer are configured according to the needs of the neural network. The number of

neurons in a layer controls the number of entities the layer can recognize and the number of layers it controls. Realizable functional complexity in these classes[37]. Figure 3 shows an example of a feedforward neural network.

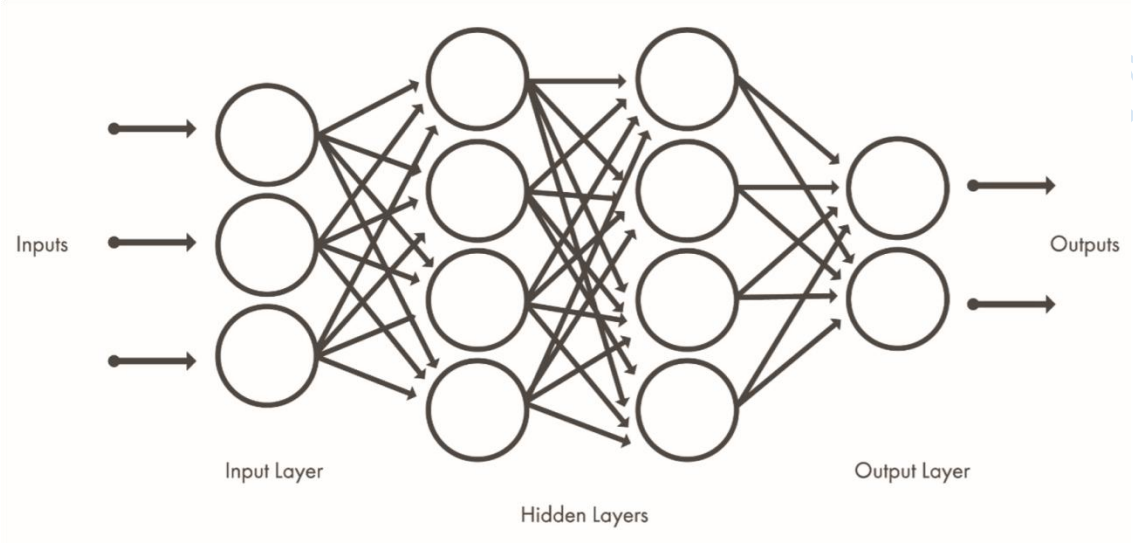

Figure 2.3. An Example of a Feedforward Artificial Neural Network.

Sources: Online (2021)

2.1.21 Overview of Cybercrime and Cybersecurity

As generations evolved, so did the definitions of cyberspace, cyberprotection, and cybercrime. Since laptop crime can include all types of crime, the definition argues that the specificity, know-how, or usage of the laptop generation needs to be emphasized. Cyberarea refers to a borderless area called the Internet. This refers to the interdependent community of additives that produce the statistics that underlie much of the media technology in the region today. Cybersecurity is a tool, policy, protection concept, protection, policy, randomized controlled trial, countermeasure, training, comfortable practice, assurance, technology that can used to protect the network environment, company assets and users. It was a gathering.

Organizational and user assets including associated computing equipment, personnel, infrastructure, applications, services, communication systems, and any statistics transmitted and/or stored stored in the network environment. Cyber Protection aims to ensure that the secure residence of the company and users' assets is reached and protected from applicable protection threats in a network environment[38]. Cyber Protection is a regulatory framework installed for the security of the cyber domain. However, as we become more dependent on cyberspace, we are certainly facing new threats. Cybercrime is a collection of crimes prepared to attack and protect any cyber area. Skilled cybercriminals and nation-states pose a threat to our economic system and national protection, among other things. Nigeria's financial strength and national defense rests on a number of essential and interdependent networks, systems, services and resources, known as cyberspace. Cyberarea has changed the way we communicate, travel, power our homes, manage the economy, and access government services. Cybersecurity is a generation, technology, and practice framework designed to protect networks, computers, packets, and records from attack, sabotage, or legitimate access. In an IT or network context, the term protection actually refers to network security[39]. Ensuring cybersecurity requires a collaborative effort by all US residents and US statistical systems. The opportunities posed by our cyber protection breaches go faster than we can tolerate. It is unrealistic to pay attention to the most practical problem of violation. This is because it leads to negligence and boom consideration for various factors of the violation. This allows us to conclude that we must attack the entire security hole of the computer.

Cybercrime is a criminal act committed using a computer and the Internet. This includes everything from illegal music file downloads to millions of dollar theft from online bank accounts. Cybercrime also includes non-monetary crimes such as creating and spreading viruses on other computers and posting sensitive business information on the Internet. Perhaps the most well-known form of cybercrime is the theft of personal information, if a criminals use the Internet to steal other users' personal information[40]. Meanwhile, the most comprehensive definition of cybercrime is illegal access (unauthorized access), illegal interception (computer data, computer system), data disruption (unauthorized damage, deletion, deterioration, modification or suppression of computer data), system failure (computer data input, transmission), Affecting computer system functionality through damage, deletion, deterioration, modification or suppression, device misuse, counterfeiting (ID-theft) and electronic fraud[41].

2.1.22 Cyber Security Goals

Below are the cybersecurity goals:

1. Help people reduce vulnerabilities in information and communication technology (ICT) systems and networks.

2. Helping individuals and organizations develop and maintain a culture of cybersecurity.

3. Cooperate with public, private and international organizations to protect cyberspace.

4. Understand current IT/cybercrime trends and develop effective solutions.

5. Availability.

6. Integrity. This may include reliability and non-repudiation.

7. Privacy.

2.1.23 Ethical Hacking and Cybercrimes in Contemporary-day Nigeria

The two major ideology in the acts of war are the knowledge of self and perceived threat “If you know the enemy and know yourself, you need not fear the result of a hundred battles" and “To protect your organization’s information, you must know yourself; that is, be familiar with the information to be protected and the systems that store, transport, and process it; and know the threats you face”[42]. Nigeria has experienced a systematic shift from the old format managing to the advanced digital age. This advancing in technology deliveries is usually accompany with a price. Importantly this development come with huge concerns such as cybercrime and electronic crime. These unexpected problems create a lot of inbalances to the system which has negative implications on both private and public sector. There is inadequate knowledge and publicity with reference to crime associated with eletronic crimes are major factors. The level of public knowledge and awareness is increasing the number of reported cyber attacks. The Cybercrime Act was only recently passed by lawmakers, but the law does not fully cover all aspects of the problem. Cybercrime is more complicated than you think. It is as attractive as any location in the country. However, as modern threats increase, Congress must continually work with investigations and informants to ensure that cybersecurity legislation is reviewed and revised at any time. Modern violations are detected.

Analyzing the events of these opportunities, most cyber attacks (based on multiple reports from various authorities across the country and beyond) provide escape clauses or attacks that give attackers access to easily selectable assets. We can conclude that it is due to a vector.

Up. Sometimes it happened by accident, most often it was caused by tinkering and human error, sometimes it was caused by a thirst for inner self and harm. Therefore, these cyber address violations are persuaded by certain unsurprising, eccentric components[43].

Reporting a comparison, one researcher found billions of naira operating in cyberspace, either through software vulnerabilities, ignorance, negligence, or direct attacks on assets aimed at causing specific harm. He added, "The overall epidemic of cyberattacks is 50% hacktivism, 40% cybercrime, 7% cybersurveillance, and 8% cyberwarfare.[44].

Similarly, the Central Bank of Nigeria recently confirmed that more than 20 billion naira were stolen due to cybercrime. This could lead to the closure of many businesses across the country and discourage many foreign investors from investing in the private and public sectors. According to the latest statistics, most of the information and resources stolen from many companies are due to the negligence and manipulation of the leaders of these companies, and 60% of former employees are held responsible for the incidents. detailed network. increase. It is believed that there are security breaches and financial disclosure of confidential information to competitors and perpetrators[45]. It is important to allow reports from Nigerian Police Fraud Units. Their latest report points out that new cybercrime techniques used by criminals have been discovered. The two most attractive are stock fraud. A stranger call you out of nowhere and tries to offer a deal at a company that didn't exist. The following is like blackmail-this type of attack focuses specifically on people in groups such as religions and ethnic groups to lure them into traps that can be fooled[46].

More importantly, NortonLifeLock's survey of 20,000 cybercrime victims in 24 countries reveals that 69% of the population is victims of cybercrime once or twice in their lifetime. became. According to the report, 14 victims are being attacked per second, which means more than one million attacks are made each day[47]. However, more than 50% were due to inadequate execution or the need for proper updates and frustration when performing regular vulnerability scans.

A recent forecast by the central bank governor during a speech at the central bank's quarterly Supreme Auditor meeting was that "cybercrime losses across all segments in 2015 were estimated at approximately $ 400- $ 550 billion. It was something like. He added that this number could reach about $ 2 trillion by the end of 2019[48].

2.1.24 Impact of Ethical Hacking and Cybercrime on the Country

1. Loss of Information

Organizational information is the most sensitive asset in all countries. The information can be anything. It is the discovery of government behavioral investigations, including deep and private content, that could fund security guards, security guards, and the best government agencies that could be particularly at risk if information were leaked. In the unlikely event that such information is maliciously sold to perpetrators or competitors, it can result in enormous losses and notoriety for the country or organization.

2. Business Disruption

The Internet is the most public platform used by almost every company in Nigeria. They are focused on their daily trading. As a result of multiple security breaches, many had to shut down their online businesses for fear of attack. Alternatively, a foreign investor declined an offer to invest in a Nigerian company because of cybercrime. As a result, many e-commerce and e-commerce businesses have suffered extreme collapse in recent years.

3. Loss of Reputation

This is one situation in which a country loses respect from investors and other international organizations. When an individual does not know what your abilities are, it becomes very difficult to do anything at this point. Reputational decline is one of the biggest cybercriminal disasters. Each year, more than 51 countries suffered a decline in reputation[49]. Similarly, in an interview with Telegraph, President Muhammad Buhari made the same statement that "the reputation of Nigeria's crimes no longer welcomes the country abroad.

4. Loss of lives

Numerous reports from the United States confirm that several murders were committed as a result of hacking medical devices[50]. The field of medicine is now evolving compared to 10 years ago. The use of computerized machines and chips is now widespread in advanced hospitals. The patient is observed with a monitoring device directly connected to the computer. In some cases, they hack the drug menu and change the drug recommended by the expert. Thus, experts are misunderstood to administer the appropriate medication to the patient. And this resulted in the death of many patients due to drug overdose[51].

2.1.25 The Dimension of Ethical Hacking and Cybercrimes

In studying the size of cybercrimes, many elements contributed to what's visible as cybercrimes. Presently, entities aren't cognizance at the outer global assaults anymore. Attacks should both come from insiders or outsiders. Insider assault is assault achieved via way of means of disenchanted or noxious worker even as outsider assault is assault achieved via way of means of a malicious individual, now no longer within the organization[52]. Insider assaults may be taken into consideration because the finest threats to any device or organization. Insiders have get right of entry to and privileges to without difficulty collide with outsiders to introduce a backdoor that would compromise the organization`s infrastructure. Also, it's far hard to discover such insider, if they're on the pinnacle of the managerial hierarchy. In addition, they're tremendous in erasing their hint as they assault. One essential aspect really well worth specifying is the reality that they are able to insert a noxious code in the device that appears purpose actual damage once they have cleared out the organization.

In other words, outsider assaults are plenty less complicated to control if as compared with insider assaults. When a valid test is performed on a device, the directors can block any destiny incidence of such attacks[53].

2.1.26 Classification of Ethical Hacking and Cyber Attack

1. Hacking

Hackers exploit operating system vulnerabilities and loopholes to corrupt data and steal sensitive information from victims' computers. This is usually done using a backdoor program installed on your computer. Many hackers try to access resources using password hacking software. Hackers can also monitor what they are doing on their computer and even import files to their computer. Hackers can unknowingly install multiple programs on your system. Such programs can also be used to steal personal information such as passwords and credit card information. You can also hack important company data to get confidential information about your company's future plans.

2. White Hats

White hat hackers or ethical hacker don't expect to harm the system or organization but they do so, authoritatively, to enter and find the vulnerabilities, providing solutions to fix them and guarantee security.

3. Black Hat

Black hat hackers or non-ethical hackers perform hacking to fulfill their selfish intentions to collect monetary benefits.

4. Grey Hat

Grey hat hackers are the combination of white and black hat hackers. They hack without any malicious intention for fun. They perform the hacking without any approval from the targeted organization.

5. Suicide Hackers

Hackers in this category have reportedly appeared recently. They may be individuals or groups that can shut down any system or infrastructure for a particular purpose. They are not afraid to be prosecuted for their actions.

6. Spy hackers

They are individuals with advanced skills and equipment, and their knowledge of hacking is great. They usually work for private organizations and governments. In particular, they are hired to spy on people's resources and personal lives.

7. Cyber terrorists

Attackers in this category are considered the most dangerous of all groups. They consist of qualified individuals or groups whose primary purpose is to destroy individuals or governmental organizations. They are highly organized and have the necessary foundations to fend off all types of attacks.

8. Cyberbullying

Cyberbullying is an intentional and electronic threat against an individual. These actions include cyber harassment.

9. Online Money Laundering

Online money laundering is the transfer of illegally obtained electronic money to conceal its source and possibly destination.

10. Website Duplication

Website Duplication: A recent trend of cybercriminals is "fake" websites abusing consumers who are new to the internet or don't know the exact web address of the legitimate business they're trying to get. access. It was an appearance. Consumers believe they are entering credit information to purchase from the target company, and instead they have unwittingly entered the information into the fraudster's personal database. Fraudsters can then use this information for their own purposes or sell it to others interested in credit card fraud.

2.1.27 Hacking and Cyber Tricks Emerging in Nigeria

1. **Beneficiary Will Scam:** Criminals send emails claiming that victims are beneficiaries named in the will of a distant relative and inherited millions USD worth of assets.
2. **Online Charity:** Another common aspect of crime in Nigeria is that scammers host charity websites seeking funding and materials for non-existent organizations. Unfortunately, many helpless people have been taken advantage of in this way.
3. **Continuation of Relative Scam:** Collect and transfer fees from different banks when you try to collect millions of dollars from a Nigerian bank owned by a deceased relative.

4. **Computer/Internet Service Time Theft**: Whiz Kids of Nigeria has developed a way to connect Cyber Cafe to the network of some ISPs in a way that is not recognized by ISPs. This allows the cafe to operate for free.
5. **Lottery Scam:** The convince users that they are actually the beneficiaries of the online lottery, which is a scam[54].

2.1.28 Challenges of Cybercrime

1. ICT Security Adviser and member of the Nigeria Cybercrime Working Group (NCWG), announced that the rate of electronic crime in Nigeria exceeds the rate of domestic internet use. He said Nigeria is the 56th out of 60 countries embracing Internet usage but third in the fraud attempt category. We are tempted to ask why there is such an upsurge of e-crime in Nigeria and what are the factors that made Nigerians so vulnerable to e-crime?
2. **National and International Law Enforcement:** Hostile parties can use internet-connected computers thousands of miles away and easily attack Nigeria's internet-connected computers as if they were outside edge. It is often difficult to identify the perpetrators of such an attack, and even when the perpetrator is identified, cross-border prosecution is a dilemma.
3. **Unemployment:** The flood of unemployment in Nigeria is at an alarming rate and increasing day by day. Bankruptcy of businesses and bankruptcy of financial institutions. The federal government has proposed a mass layoff of civil servants. Companies are also embarking on mass layoffs. Financial institutions have placed

unreasonable age barriers on those eligible to apply for jobs and have embarked on mass layoffs based on ad hoc decisions.

4. **Poverty Rate:** Nigeria is considered a third world country in the world. The poverty rate continues to rise. The rich get richer and the poor get poorer. Insufficient basic equipment and epileptic power supplies wiped out small industries.
5. **Corruption:** Nigeria ranks third among the most corrupt countries in the world. Until 1999, corruption was considered a way of life in Nigeria.
6. **Absence of Standards and Centralized Control in the Country:** Charles Emeruwa, adviser to the Nigerian Cybercrime Working Group (NCCWG), argues that the absence of regulations, standards and IT security and protection laws impede true e-commerce. Foreign Direct Investment (FDI) and outsourcing encourage the misuse and abuse of computers.
7. **Lack of Infrastructure:** Updated information and communication technology equipment for appropriate surveillance and arrest.
8. **Lack of Functional Country Databases:** National databases can be used to track down the perpetrators of these ills by looking at past personal records and tracking transfers. their movements.
9. **Cyber Cafe Dissemination:** Many entrepreneurs act as a blissful paradise for entrepreneurs to practice their actions through night browsing services that Syndicate provides to potential customers without guidance or supervision to achieve their goals. It was started.
10. **Internet Porousity:** The Internet is free for everyone and is not centrally controlled. Therefore, the current state of anarchy is as stack.

2.2 Theoretical Review

Theoretical framework of this research is a structure that can hold or support the theory of the research paper. It represents a theory that explains why the problem under study exists. Therefore, the theoretical framework is only the theory underpinning the conduct of the research. The theoretical framework guides the investigation and determines the statistical relationship between what is measured and what is searched.

The theories used in this study are structural functional theory, Marxian theory, frequent activity theory, and technologically enhanced crime theory.

2.2.1 Structural Functionalist Theory

The central finding of structural functionalist theory is that crime and deviation are a necessary part of social organization. She believes that society is an agency, a system of parts that perform functions together for the common effectiveness and efficiency of society. Structural functionalism is a theory of consensus that society is built on order, reciprocal relationships, and a balance between parts to keep the whole running smoothly. The theory that shared norms and values are the foundation of society, focuses on social order based on tacit consensus between groups and organizations, and considers social change to be slow. slow and orderly.

In the mid-1930s, Merton understood and wrote about crime and deviance as a response to the failure to achieve social goals. This is known as the "safety theory" of crime, as Merton emphasizes tensions or tensions between:

1. The cultural goals of the society and

2. Legal or institutionalized means of achieving those ends. The relevance of this theory to this study is to give us the idea that crime and deviance are not a matter of a few bad apples. It is a necessary condition for a "good" social life. The theory is that to control crime, governments must legislate and create an institutional framework to enforce law, order, and cybersecurity in Nigeria.

2.2.2 Marxian Theory

The main lesson of the theory is that crime is a natural consequence of capitalism and that society is ever-changing in response to social inequality and social conflict. Capitalism as an economic system is based on private ownership of property for personal gain, rather than the promotion of collective welfare. The theory that capitalism itself is a crime and continues to do it. Based in large part on economic oppression and exploitation, it creates a highly competitive world of greed, violence and corruption.

It provides a very quick interpretation of Marxist ideas about crime and deviation. Bonger shared with Marx himself the belief that humanity is inherently altruistic and uncompetitive. Bonger suggests that capitalism itself is a form of economic organization that makes humanity greedy and selfish[55].

Consistent with Bonger's argument, he argues that under capitalism, the law is used to oppress the working class. He suggests that what we now consider "criminal" will not go away until capitalism itself is gone. He argues that under socialism there would be no greed or profitability. In addition, the ruling class does not exist to use the law as a weapon to deviate from the activities of the working class or to define it as a criminal[56].

Marxian's theory is relevant to this research because it provides an important insight into why people, especially unemployed youth, are criminals. Given the high levels of political and economic instability and corruption in Nigeria, it is not surprising that cybercrime is widespread. For the oppression, exploitation and alienation of the majority for the benefit of the elite. Some disadvantaged citizens, the majority, have chosen alternatives to survive. These alternatives include prostitution, armed robbery, etc. It turns out that crime in Nigeria is, among other things, influenced by extreme poverty, relative social deprivation, rampant corruption, excessive greed and materialism.

2.2.3 Routine Activity Theory

They argued that there must be three conditions for a crime to be fulfilled, which are: Lack of motivated abusers, good targets, and competent guardians[57].

The theory that crime is normal and depends on the choices available. If the target is not properly protected and the reward is worthy, crime will occur. Crimes didn't require hard-line criminals, super-robbers, convicted serious offenders, or malicious people. Crime

requires only one chance. When committing a crime, he states that the three elements must be in the same room at the same time, which are:

1. Appropriate targets are available

2. There are no appropriate guardians to prevent crime

3. There are likely and motivated perpetrators.

This theory is relevant to this study in that it provides a deeper understanding of why people are involved in cybercrime. Cybercrime has to do with the effectiveness of indirect guardians, as such a motive for such a crime. In addition, the Global Information Infrastructure (GII) is open and over kill, and Internet mechanisms are designed to send data rather than inspect it.

2.2.4 Technology-enhanced Crime Theory

The main idea of the theory is that it combines some kind of criminological theory to help society better understand why crime has co-evolved with computer and telecommunications technology to become one of the types of crime. most complex and more difficult to prevent, investigate, and control. It has been revealed that relatively complex crime is initially quite difficult to understand and maintain, and that there is constant competition between criminals and law enforcement agencies for technological advantages[58]. When criminals do something new and innovative, law enforcement must catch up to prevent, control, deter and deter new forms of crime.

Technology-based crime theories are said to include[59]:

1. Crimes of directly infringing on computers and computer systems.

2. Activities of this type are often referred to as tech crimes, computer crimes or cybercrimes.
3. Using technology to commit or promote traditional crimes.
4. Crime, fraud, harassment and other crimes can be facilitated by using technology that presents unique challenges to ancient crimes.

Theory provides a framework for understanding all forms of crime, especially those evolving with the invention and innovation of computer and telecommunications technology. This theory relates to understanding new forms of cybercrime, cross-border crime, and current threats from terrorist networks that go against traditional criminal justice systems and security measures to prevent and control crime.

This theory is relevant to this study as it provides insights into the new tools and techniques used by cybercriminals. In other words, it is a transition from a simple crime committed by simple means to a complex crime committed by complex means. It also helps us understand new forms of deviations, social abuses, or crimes that are being committed through the innovative use of technology.

2.3 Review of Related Works

A study on Data Mining Techniques and Algorithm. The author proposed the building of Classifiers for sentiment analysis which consists of different machine learning classifiers. Suspicious behaviours are categorized under groups such as terrorist activity, financial laundering, hacking, sexual or racial harassment[10].

A study on Suspicious Pattern Detection (SPD) Algorithm for the identification of suspected cyber threat in instant chat messenger available on Social Networking Websites and Instant Messengers. The proposed framework has considered the Ontology based Information Extraction technique (OBIE) with a pre-defined knowledge base data mining approach of Association Rule Mining (ARM). The proposed concept involves three major steps as mentioned: (a) word extraction from unstructured text (b) e-crime monitoring system program (c) SPD algorithm. The proposed concept has been tested for the Global Terrorist Database (GTD). The proposed concept has been compared with other Instant Messengers, Mobile Phone Apps and Social Networking Sites based on the ability to detect suspicious information during online chats. As per considered parameters, proposed concept shows efficient results[30].

Most of the data of online forums are stored in text format as noted by, therefore the present work makes use of only text format of suspected postings as evidence for investigation. But there are many suspicious users such as spammers, fraudsters, and other types of attackers that use the latest technology for the criminal activities. The researchers discovered the existence of tools and methods for the recognition of suspicious information available on internet in the form of user comments or views. The Author proposed a framework that use integrated approach of Support Vector Machine (SVM) and Particle Swarm Optimization (PSO). SVM is a statistical learning based data mining approach and PSO is swarm intelligence based concept considered to optimize the parameters of SVM.

A review was carried out on the use of Term Frequency–Inverse Document Frequency (TF-IDF) for document s classification. On the basis of these features, logistic regression and linear support vector machine classifiers were used; both classifiers train rapidly, require little computation to analyse a document, and provide an output score proportional to the probability that the input document contains cyber content. Results was in the form of Detection Error Tradeoff (DET) curves that show how false-alarm and miss probabilities vary as the threshold on the classifier's output probability varies as plotted on normal deviate scales. Logistic regression classifier performs better than the keyword system and the logistic regression classifier passes through the performance target region, meaning it misses less than 10% of cyber documents with a false-alarm rate of less than 1%.

A study used statistical corpus based data mining approach for the detection of suspicious activities on online forums. Authors have presented the work on textual data of online forums. The complete process of suspicious information extraction has been explained. After the pre-processing steps of stop words removal and stemming process with Brute Force algorithm, authors have used the matching algorithm for the suspicious keyword recognition. Finally, authors have used the keyword spotting techniques, leaning based method and hybrid of defined approaches for the overall recognition of suspicious human activity.

A finding was also reported in a study that introduced the novel algorithm of CrossSpot to spot the suspicious information and fraud deviations. Authors have used the metric based approach to define the suspiciousness of a block of information from multimodel data. Initially, authors have performed the experiment with the concept on the synthetic data which

is based on Erdos-Renyi-Poisson model. Then, twitter based Hashtag hijacking dataset has been used for experimentation. The proposed concept of CrossSpot has been compared with the approach of Singular Value Decomposition (SVD) and High Order SVD (HOSVD). The proposed concept of CrossSpot shows efficient results for the experimentation with the presentation in terms of F1 Score.

This researcher proposed an approach to differentiate the twitter data as an actual information and rumor. Authors have extracted the twitter data for some particular topic with the help of Hashtag functions. For the validation of concept for any particular information, data of some well-known news channels has been considered and evaluated the results with semantic and sentiment analysis of tweets. For this proposed concept, authors have also presented a prototype "The Twitter Grapevine" to target the rumors specifically for Indian domains. The overall results have been evaluated based on the accuracy analysis initially for digital India & facebook.org rumors and then for KeralaHouse & Beef Rumor topics. In these results, favourable and unfavourable result predictions have been evaluated. Accuracy results for the later experiments are much lower than the former one of 76.99%.

A study on the approach of crime detection and criminal identification (CDCI) using data mining approach. Authors have used the Indian dataset if criminal acts like Delhi rape cases, national crime records, committee to protest journalists, crime alerts etc. Further, data has distributed into 35 major categories with their attributes. The results have been evaluated for the seven major Indian cities as mentioned: Bangalore, Hyderabad, Jaipur, Pune, Mumbai, Kolkata and Delhi. For the simulation of results, the authors have used Java based Netbeans

platform for the identification and prediction of crime and criminal activities. Also, WEKA tool has been used for the verification of crime activities. KNN approach has been considered for the criminal identification and Google maps have been embedded to enhance the k-means clustering. A graphical user interface has been designed for this CDCI approach and efficient accuracy results have been achieved for considered concept.

A study on a similarity distance evaluation based approach to differentiate the suspicious content from the other authentic content/posts/blogs. The dataset of twitter text corpus has been used for the analysis. The considered concept is based on the evaluation of similarity index by comparing the social database. There are basically three steps of proposed concept as mentioned: text corpus, corpus processing and classification process using similarity approach. The concept is evaluated in terms of similarity distance index which leads to more execution time and lesser precision rate. So, there is need to improve the precision rates and execution time.

A review for the existing concepts of crime data mining techniques for the detection of suspicious information on the web. The major challenge for the detection of suspicious concept is the day by day increase in volume of cyber data and increasing network traffic. There is the availability of various data formats like audio, video and textual data. In this research work, a theoretical review for the different mining concepts like data mining, web mining, crime data mining has discussed. On the basis of textual data mining concepts for the detection of crime activities are sequential pattern mining, classification, association rule mining and clustering etc.

Social media offers a variety of avenues through which we can communicate with people. In fact, social media is known to have been used widely in educational field also. Over the last 30 years the nature of communication has undergone a substantial change and it is still changing. Email has had a profound effect on the way people keep in touch. Communications are shorter and more frequent than when letters were the norm and response time has greatly diminished. Instant messaging has created another method of interaction, one where the length of messages is shorter and the style of the interaction is more conversational. Broadcast technologies like Twitter transform these short bursts of communication from one-on-one conversations to little news (or trivia) programs: which we can tune in, whenever we want an update or have something to say. The traditional data mining techniques just classify the patterns in structured data for example, classification and prediction, association analysis, outlier analysis and cluster analysis. On the other hand, the newer techniques identify patterns from unstructured and structured data[11].

2.4 Summary of Literature Reviewed

This chapter involves the review of related literature to this study. The basic concepyt of this study were reviewed based on the past scholarly publications. The study adopted Structural functionallis, Theory, Marxian Theory, Routine Acitivity Theory and the Theory of Technology-enabled Crime. Emperical studies that were related to this study were succinctly reviewed to include the objective, methodology, major findings, and conciusion where applicable.

Endnotes

1. T. Hanley, J. Prescott & K.U. Gomez, A Systematic Review Exploring How Young People Use Online Forums for Support Around Mental Health Issues. *Journal of Mental Health*, *28*(5), 2019, pp.566-576.

2. T. Roelen-Blasberg, J. Habel & M. Klarmann, Automated Inference of Product Attributes and Their Importance from User-Generated Content: Can We Replace Traditional Market Research. *International Journal of Research in Marketing*, 2022.

3. J.W. Morris, Hearing The Past: The Sonic Web from MIDI to Music Streaming. *The SAGE Handbook of Web History. Londres: SAGE Publications,* 2019, pp.491-504.

4. M. Leithner & D.E. Simos, CHIEv: Concurrent Hybrid Analysis for Crawling and Modeling of Web Applications. *ACM SIGAPP Applied Computing Review,* 21(1), 2021, pp.5-23.

5. D. Hoogeveen, Real and Misflagged Duplicate Question Detection in Community Question-Answering *(Doctoral Dissertation),* 2018.

6. V. Creelman, "Thank You for Reaching Out:" Brand Relationship Management and The Conversational Human Voice of Customer Care in Online Service Encounters. *Discourse, Context & Media, 46*, 2022, p.100572.

7. L. Canter, The Misconception of Online Comment Threads: Content and Control On Local Newspaper Websites. *Journalism Practice,* 7(5), 2013, pp.604-619.

8. V. Terziev, Shared Knowledge and Presentation of the Research Results in Specialised Scientific Publications*. Available at SSRN,* 2022.

9. M.A. Hoque & C. Davidson, Design and Implementation of an IoT-Based Smart Home Security System. *Int. J. Networked Distributed Comput.,* 7(2), 2019, pp.85-92.

10. J. Li, Q. Xu, R. Cuomo, V. Purushothaman & T. Mackey, Data Mining and Content Analysis of the Chinese Social Media Platform Weibo During the Early COVID-19 Outbreak: Retrospective Observational Infoveillance Study. *JMIR Public Health and Surveillance,* 6(2), 2020, p.e18700.

11. A.M. Hasan, T.H. Rassem, N.M. Noor & A.M. Hasan, September. A Semantic Taxonomy for Weighting Assumptions to Reduce Feature Selection from Social Media and Forum Posts. *International Conference of Reliable Information and Communication Technology*, 2019, (pp. 407-419). Springer, Cham.

12. O. Almatrafi & A. Johri, Systematic Review of Discussion Forums in Massive Open Online Courses (MOOCs). *IEEE Transactions on Learning Technologies*, 12(3), 2018, pp.413-428.

13. A.S. Slavko, V.M. Zavhorodnia & N.Y.A. Shevchenko, Protection of One's Honor, Dignity, and Business Reputation on Social Networks: Issues and Ways to Resolve Them, 2020.

14. F.B. Osang & V. Nwaocha, Bridging The Distance in Open and Distance Learning: Developing Student-Student and Student-Lecturer Collaborative Forum. *Lautech Journal of Engineering and Technology,* 12(1), 2018, pp.97-112.

15. A.D. Nobari, M.H.K.M. Sarraf, M. Neshati & F.E. Daneshvar, Characteristics of Viral Messages On Telegram; The World's Largest Hybrid Public and Private Messenger. *Expert Systems with Applications,* 168, 2021, p.114303.

16. N. Kwon, H. Deshpande, M.K. Hasan, A. Darnal & J. Kim, Multi-ttach: Techniques to Enhance Multi-material Attachments in Low-cost FDM 3D Printing. *Symposium on Computational Fabrication,* 2021, (pp. 1-16).

17. C. Crisci, B. Ghattas & G. Perera, A Review of Supervised Machine Learning Algorithms and Their Applications to Ecological Data. *Ecological Modelling,* 240, 2012, pp.113-122.

18. Y.K. Wong, Advanced Deep Learning Approach and Applications. *International Journal of Information Technology (IJIT),* 7(5), 2021.

19. N. Dutta, N. Jadav, S. Tanwar, H.K.D. Sarma & E. Pricop, Intrusion Detection Systems Fundamentals. *Cyber Security: Issues and Current Trends,* 2022, (pp. 101-127). Springer, Singapore.

20. T. Saba, T. Sadad, A. Rehman, Z. Mehmood & Q. Javaid, Intrusion Detection System Through Advance Machine Learning for The Internet of Things Networks. *IT Professional,* 23(2), 2021, pp.58-64.

21. A.A. Ojugo & R.E. Yoro, Forging A Deep Learning Neural Network Intrusion Detection Framework to Curb the Distributed Denial of Service Attack. *International Journal of Electrical and Computer Engineering, 11*(2), 2021, p.1498.

22. G. Nguyen, S. Dlugolinsky, M. Bobák, V. Tran, A. Lopez Garcia, I. Heredia, P. Malík & L. Hluchý, Machine Learning and Deep Learning Frameworks and Libraries for Large-Scale Data Mining: A Survey. *Artificial Intelligence Review,* 52(1), 2019, pp.77-124.

23. B.R. Bhamare & J. Prabhu, A Supervised Scheme for Aspect Extraction in Sentiment Analysis Using the Hybrid Feature Set of Word Dependency Relations and Lemmas. *PeerJ Computer Science,* 7, 2021, p.e347.

24. R. Sahani, C. Rout, J. Chandrakanta Badajena, A.K. Jena & H. Das, Classification of Intrusion Detection Using Data Mining Techniques. *Progress In Computing, Analytics and Networking,* 2018, (pp. 753-764). Springer, Singapore.

25. Y. Ming, H. Qu & E. Bertini, Rulematrix: Visualizing and Understanding Classifiers with Rules. *IEEE Transactions On Visualization and Computer Graphics,* 25(1), 2018, pp.342-352.

26. I. Obeidat, N. Hamadneh, M. Alkasassbeh, M. Almseidin & M. AlZubi, Intensive Pre-Processing of Kdd Cup 99 For Network Intrusion Classification Using Machine Learning Techniques, 2019.

27. A.Y. Xue, J. Qi, X. Xie, R. Zhang, J. Huang & Y. Li, Solving The Data Sparsity Problem in Destination Prediction. *The VLDB Journal,* 24(2), 2015, pp.219-243.

28. N.B. Thylstrup, The Ethics and Politics of Data Sets in The Age of Machine Learning: Deleting Traces and Encountering Remains. *Media, Culture & Society,* 2022, p.01634437211060226.

29. L. Dhanabal & S.P. Shantharajah, A Study On NSL-KDD Dataset for Intrusion Detection System Based On Classification Algorithms. *International Journal of Advanced Research in Computer and Communication Engineering,* 4(6), 2015, pp.446-452.

30. V. Kumar, A.K. Das & D. Sinha, Statistical Analysis of The UNSW-NB15 Dataset for Intrusion Detection. *Computational Intelligence in Pattern Recognition,* 2020, (pp. 279-294). Springer, Singapore.

31. L. Zhiqiang, G. Mohi-Ud-Din, L. Bing, L. Jianchao, Z. Ye & L. Zhijun, Modeling Network Intrusion Detection System Using Feed-Forward Neural Network Using Unsw-Nb15 Dataset. *IEEE 7th International Conference on Smart Energy Grid Engineering (SEGE),* 2019, (pp. 299-303). IEEE.

32. Q.V. Dang, Studying Machine Learning Techniques for Intrusion Detection Systems. *International conference on future data and security engineering,* 2019, (pp. 411-426). Springer, Cham.

33. E.H. Houssein, M. Kilany & A.E. Hassanien, ECG Signals Classification: A Review. *International Journal of Intelligent Engineering Informatics,* 5(4), 2017, pp.376-396.

34. H.H. Patel & P. Prajapati, Study and Analysis of Decision Tree Based Classification Algorithms. *International Journal of Computer Sciences and Engineering,* 6(10), 2018, pp.74-78.

35. X. Zhou, P. Lu, Z. Zheng, D. Tolliver & A. Keramati, Accident Prediction Accuracy Assessment for Highway-Rail Grade Crossings Using Random Forest Algorithm Compared with Decision Tree. *Reliability Engineering & System Safety,* 200, 2020, p.106931.

36. R.C. Chen, C. Dewi, S.W. Huang & R.E. Caraka, Selecting Critical Features for Data Classification Based On Machine Learning Methods. *Journal of Big Data,* 7(1), 2020, pp.1-26.

37. N. Zhao & J. Lu, Review of Neural Network Algorithm and Its Application in Temperature Control of Distillation Tower. *Journal of Engineering Research and Reports,* 20(4), 2021, pp.50-61.

38. S. Pandey, R.K. Singh, A. Gunasekaran & A. Kaushik, Cyber Security Risks in Globalized Supply Chains: Conceptual Framework. *Journal of Global Operations and Strategic Sourcing,* 2020.

39. H. Lin, Z. Yan, Y. Chen & L. Zhang, A Survey On Network Security-Related Data Collection Technologies. *IEEE Access,* 6, 2018, pp.18345-18365.

40. S. Batra, M. Gupta, J. Singh, D. Srivastava & I. Aggarwal, An Empirical Study of Cybercrime and Its Preventions. *Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC),* 2020, (pp. 42-46). IEEE.

41. G.Y. Koi-Akrofi, J. Koi-Akrofi, D.A. Odai & E.O. Twum, Global Telecommunications Fraud Trend Analysis. *International Journal of Innovation and Applied Studies,* 25(3), 2019, pp.940-947.

42. M.H. Abdul-Rahim, Assessment of The Perceptions On the Effects of Cybercrime On Senior High Students in Tamale Metropolis: A Case Study of Vitting Senior High School *(Doctoral Dissertation),* 2021.

43. B.M. Thuraisingham, Can AI Be for Good in The Midst of Cyber Attacks and Privacy Violations? A Position Paper. *Proceedings of the Tenth ACM Conference on Data and Application Security and Privacy,* 2020, (pp. 1-4).

44. K.M. Caramancion, Y. Li, E. Dubois & E.S. Jung, The Missing Case of Disinformation from The Cybersecurity Risk Continuum: A Comparative Assessment of Disinformation with Other Cyber Threats. *Data,* 7(4), 2022, p.49.

45. F. Schlackl, N. Link & H. Hoehle, Antecedents and Consequences of Data Breaches: A Systematic Review. *Information & Management,* 2022, p.103638.

46. I. Beutel, O. Kirschler & S. Kokott, How Do Fake News and Hate Speech Affect Political Discussion and Target Persons and How Can They Be Detected? *Central and Eastern European eDem and eGov Days,* 342, 2022, pp.37-81.

47. Y. Jia, F. Zhong, A. Alrawais, B. Gong & X. Cheng, Flowguard: An Intelligent Edge Defense Mechanism Against IoT DDoS Attacks. *IEEE Internet of Things Journal,* 7(10), 2020, pp.9552-9562.

48. R.M.F. Esteves, Financial Digitalization: Risk or Opportunity *(Doctoral dissertation),* 2021.

49. K. Daugirdas, Reputation as A Disciplinarian of International Organizations. *American Journal of International Law,* 113(2), 2019, pp.221-271.

50. K. Wellington, Cyberattacks On Medical Devices and Hospital Networks: Legal Gaps and Regulatory Solutions. *Santa Clara High Tech. LJ,* 30, 2013, p.139.

51. T.C. Davis, M.S. Wolf, P.F. Bass III, J.A. Thompson, H.H. Tilson, M. Neuberger & R.M. Parker, Literacy and Misunderstanding Prescription Drug Labels. *Annals of Internal Medicine,* 145 (12), 2006, pp.887-894.

52. F. Gandhi, D. Pansaniya & S. Naik, Ethical Hacking: Types of Hackers, Cyber Attacks and Security. *International Research Journal of Innovations in Engineering and Technology,* 6(1), 2022, p.28.

53. L.Á. Almeida, Data Model Classification Based On Machine Learning Techniques for Detection of Anomalous Traffic. *Cartagena de Indias,* 2019.

54. T. Dam, L.D. Klausner, D. Buhov & S. Schrittwieser, Large-Scale Analysis of Pop-Up Scam On Typosquatting Urls. *Proceedings of the 14th International Conference on Availability, Reliability and Security,* 2019, (pp. 1-9).

55. J.W. Messerschmidt, From Marx to Bonger: Socialist Writings On Women, Gender, And Crime. *Sociological Inquiry,* 58(4), 1988, pp.378-392.

56. J. Young, Working-Class Criminology. *Critical Criminology (Routledge Revivals),* 2013, (pp. 79-110). Routledge.

57. H. Lee & K.S. Choi, Interrelationship Between Bitcoin, Ransomware, And Terrorist Activities: Criminal Opportunity Assessment Via Cyber-Routine Activities Theoretical Framework. *Victims & Offenders,* 16(3), 2021, pp.363-384.

58. R.A. Berk, Artificial Intelligence, Predictive Policing, And Risk Assessment for Law Enforcement. *Annual Review of Criminology,* 4, 2021, pp.209-237.

59. T. Holt & A. Bossler, Cybercrime in progress: Theory and prevention of technology-enabled offenses. *Routledge,* 2015.

Chapter Three

Methodology

This Chapter focuses on the methodology used to conduct the study. It discusses the techniques and methods used to fulfill the study's objective. The following section discusses the research techniques and programming used to create a monitoring suspicious discussions system.

3.1 Research Approach

Research Approach can be classified into three types; Qualitative, Quantitative and Mixed Method.

Quatitative Approach: This be decribed as the descriptive and conceptual findings collected through questionnaires, interviews, or observation. It deals with words and meanings and offers a dynamic approach to research, the researcher has an opportunity to foolow up on answers given by respondants in real time.[1].

Qualitative Approach: It is an interpretative approach that seeks to gain insight into the specific meanings and experienced behaviors of specific social phenomena through the subjective experiences of participants[2].

Mixed methods Approach: This is aresearch methodology that combines and integrates qualitative and quantitative research methods into one research study[3]. It involves collecting and analyzing qualitative and quantitative data to understand a phenomenon better and answer the research questions.

This study is based on quantitative method of research which is the process of collecting and analysing numerical data. Also to find pattern and averages, make predictions, test casual relationships, and generalize results to wider populations.

3.2 Requirement Specification

The requirement specification to achieve the success of this system required both hardware and software tools. The hardware tools are those physical electronic devices while the software tools are the written instructions in form of programs.

3.2.1 Software Implementation Tools

Anaconda is an open-source distribution for python. It is used for data science, machine learning, deep learning, etc. With the availability of more than 300 libraries for data science, it becomes fairly optimal for any programmer to work on anaconda for data science.

Jupyter is an open source web application that you can use to create and share documents that contain live code, equations, visualizations, and text.

My Structured Query Language (MySQL): MySQL is a database system used for developing web-based software applications. MySQL is an Oracle-backed open source Relational Database Management System (RDBMS) based on Structured Query Language (SQL). MySQL is fast, reliable, and flexible and easy to use.

MySQL runs on virtually any platform, including Linux, UNIX, and Windows. Although it can be used in a variety of applications, MySQL is most commonly associated with web applications and online publishing. MySQL is an important component of the open source enterprise stack called LAMP. LAMP is a web development platform that uses Linux as the operating system, Apache as the web server, MySQL as the relational database management system, and PHP as the object-oriented scripting language. Originally conceived by the Swedish company MySQL AB, MySQL was acquired by Sun Microsystems in 2008 and by Oracle when it acquired Sun in 2010. Developers can use MySQL under the GNU General Public License (GPL), but companies must purchase a commercial license from Oracle. Today, MySQL is the RDBMS behind many of the world's best websites and myriad web-based business and consumer applications.

PHP is short for Hypertext Preprocessor. It is a widely-used open source general purpose language that is suited for web and web application development. It can be embedded into HTML (The PHP Group, 2000). PHP is focused on server-side scripting, command line scripting and writing desktop applications. It also runs on all operating systems such as

Windows, Linux and MAC. PHP supports wide range of databases including MySQL, ODBC and PDO.

JavaScript will handle web page requests by sending client-side scripts to the browser. Bootstrap is the frame work that would be used to develop responsive and interactive web app.

3.2.2 Hardware Requirements

The basic hardware requirements are:

Dual Core Processor: The processor is the logic circuitry that responds and processes the basic instruction that drive the computer. Its primary functions are fetch, decode, execute and write back. For the development of the monitoring system the least processor needed is a dual core processor, anything less would cause the system to run very slow.

4GB RAM: The RAM (Read-Access Memory) is a volatile Memory in the computer system that store data and machine code currently being used. A random-access memory device allows data items to be read or written in almost the same amount of time irrespective of physical location of data inside the memory. RAM is measured both in size and speed. RAM size determine how much temporary data the computer can store and how fast it runs. For the development of this system, a RAM size and speed less than 4gigabyte will result in a slow and tiresome performance.

1TB Hard Disk (HDD): The hard disk is the main, and usually largest, data storage hardware device in a computer. The operating system, software title, and most files are stored in the hard disk drive. The least expected size of storage for this system to work effectively is 1Terabytes or more.

Moderm: Modem is short for "Modulator-Demodulator." It is a hardware component that allows a computer or another device, such as a router or switch, to connect to the Internet. It converts or "modulates" an analog signal from a telephone or cable wire to digital data (1s and 0s) that a computer can recognize.

3.2.3 System Algorithm

Step 1: Input

Step 2: Review unstructured training datasets available on online sites and instant messengers

Step 3: Remove the numeric keywords from the dataset

Step 4: Preprocess the data by applying the start and delete outages and keyword tokenization features

Step 5: Develop an array of suffixes that must be identified and removed to obtain a root word

Step 6: Boolean = Checks if the word to be stemmed exists in the dictionary

Step 7: If boolean = True then Stemming is not needed. Other Root Word = get Root Word

Step 8: Check if the suffix exists in the suffix table expanded in Step 5

Else

Goto next word

Step 9: Remove word from stop function

Step 10: Apply NLP for Feature Extraction

Step 11: Apply SVM for the detection and classification of suspicious keywords and compare them to the keyword list in the expert dataset.

SVM = SVMtrain (suspicious_level, supicious_type)

Where

SVMtrain is an SVM training function.

Step 12: Repeat the iteration steps and Collate keywords against expert datasets to get results for suspicious activity under consideration

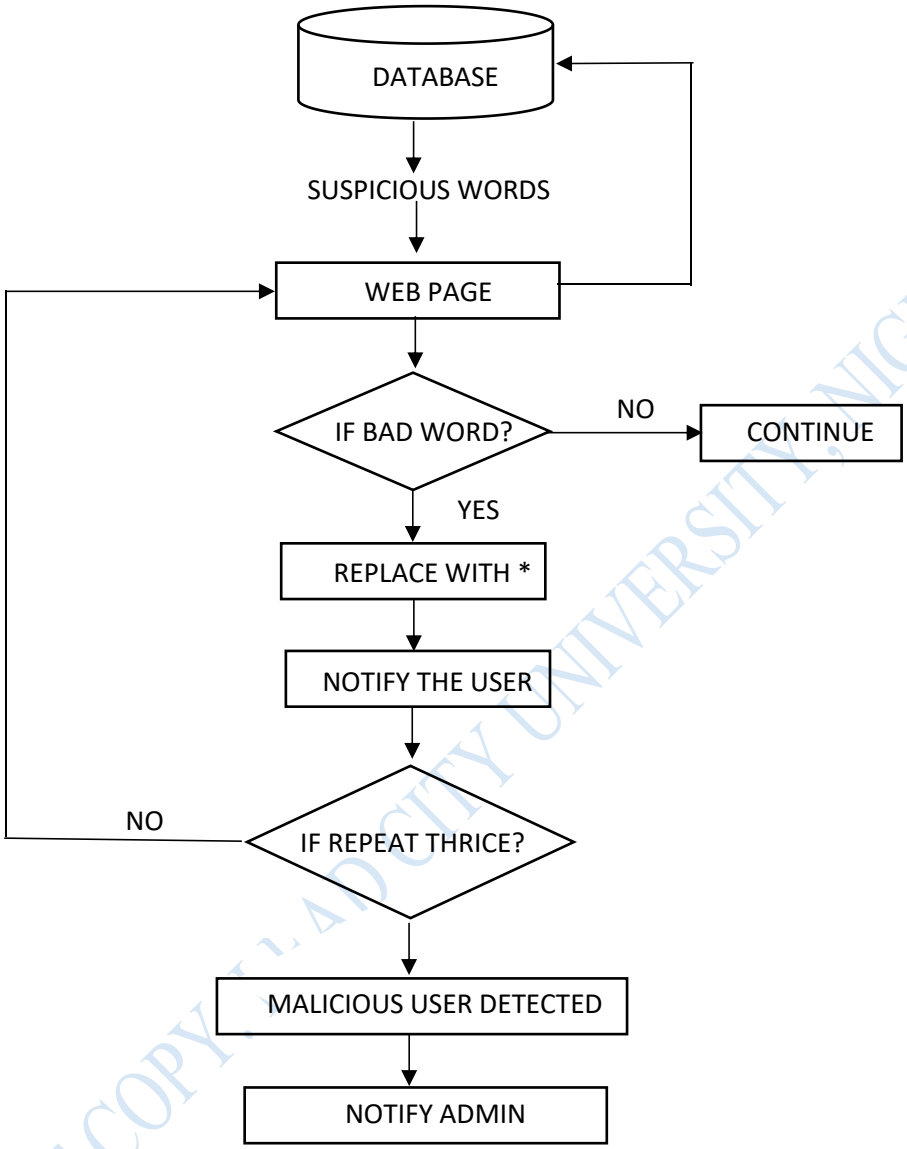

Figure 3.1 Architectural Design to Monitor Suspicious Discussions.

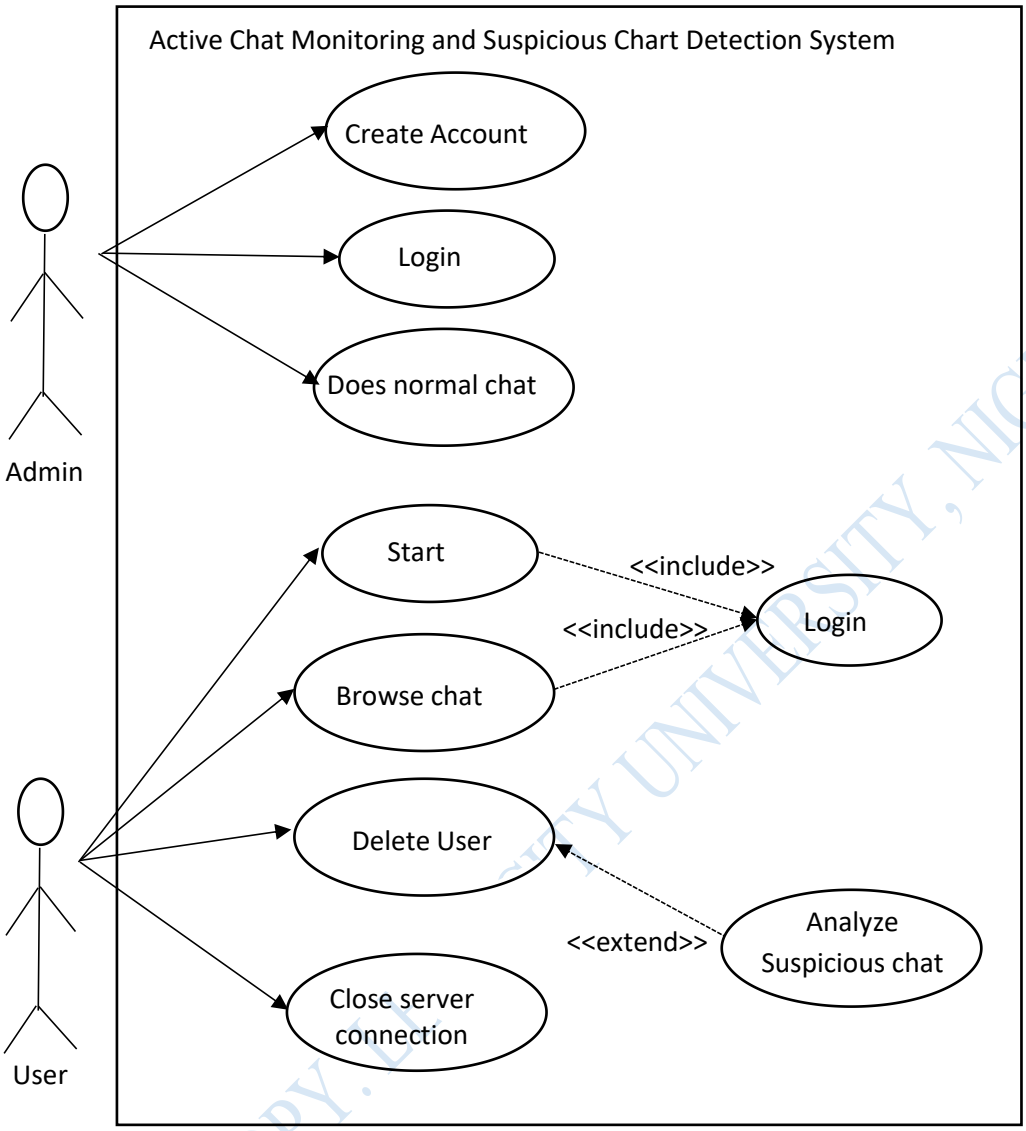

Figure 3.2: Use Case Diagram of the System.

The suspicious chat detection and chat monitoring use case diagrams show that both administrators and users need to log in to the system. After registration, users can only talk to other registered users. Meanwhile, administrators can perform a variety of tasks, such as monitoring user conversations and browsing and updating databases that record suspicious

terms. If a user uses a suspicious term, the administrator will be notified and given the option to remove the user.

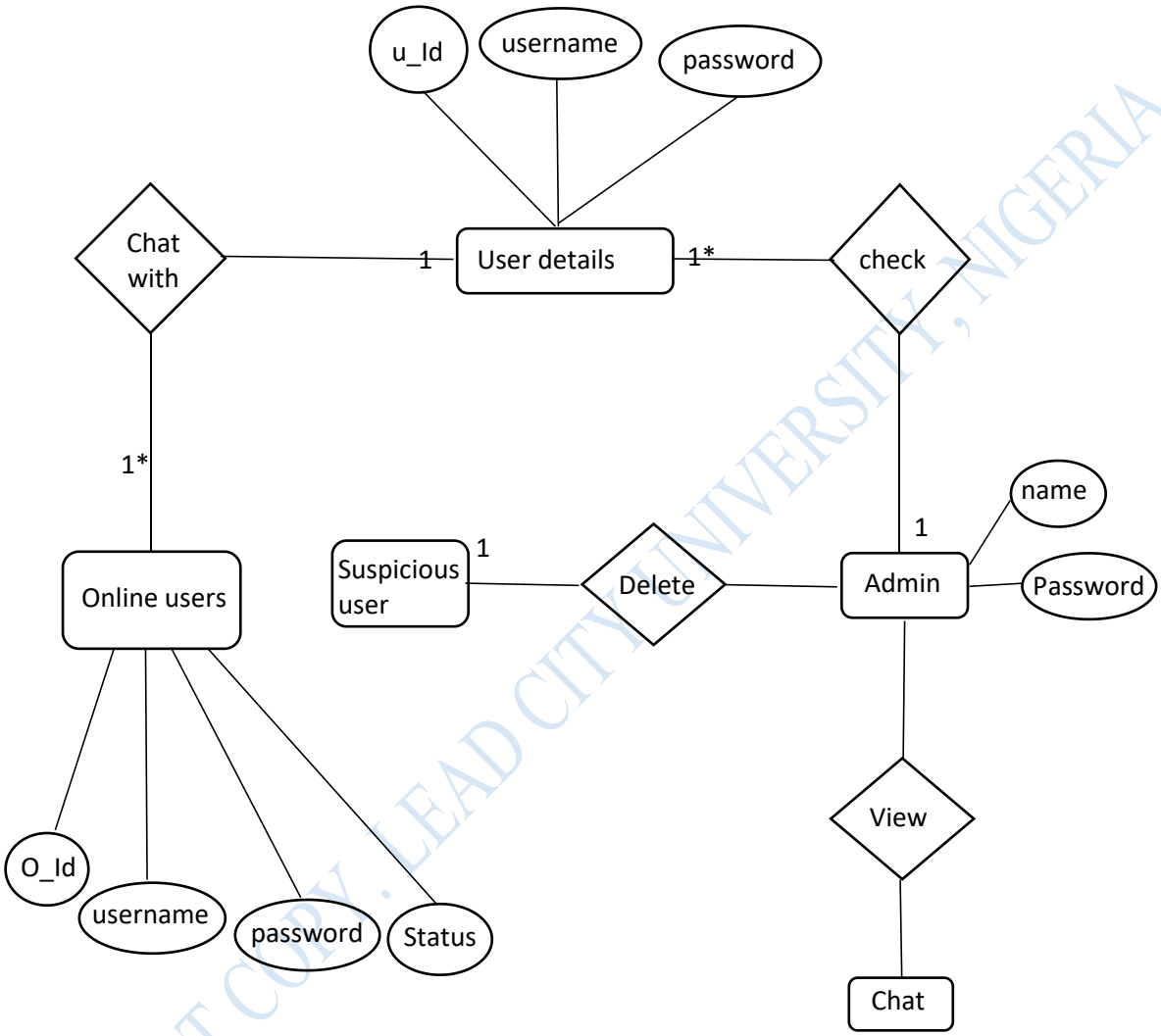

Figure 3.3: ER Diagram of Conversation Monitoring and Suspicious Conversation Detection

Entities in the ER diagram above, like user information, administrators, and online users are linked by ID. As shown in the figure, users can communicate online with other users. Users can communicate online with many people at the same time or with one person privately. If a user's chat detects suspicious text, the admin has the option to delete the user's chat.

3.3 System Design

This system will be implemented using a deep learning technique to monitor social media and discussion forums for suspicious feedback and comments.

Discussion forums allow you to reach a large number of people almost instantly. Millions of people share their views and ideas about politics and religion, and some deliberately offend religious or racist sentiment by posting malicious messages. Therefore, it is important to monitor the posts in these forums.

This thesis uses data collection from various online forums. This data is then transferred to the CSV file. On the other side of this method, users receive their account and access data on their website. You need to log in to this system to start a discussion on any topic, and whenever any of the suspicious keyword is detected, the administrator will be notified, even the user is warned about his activity.

The integrated approach of NLP and SVM is considered to detect suspicious topics in online forums. In this integrated approach, a support vector machine is a statistical learning concept used as a classification and regression model to distinguish keywords based on suspicious

activity from real information. NLP refers to the branch of computer science—and more specifically, the branch of artificial intelligence or AI—concerned with giving computers the ability to understand text and spoken words in much the same way human beings can.

The technique uses SVM and NLP implemented using Python programming languag.

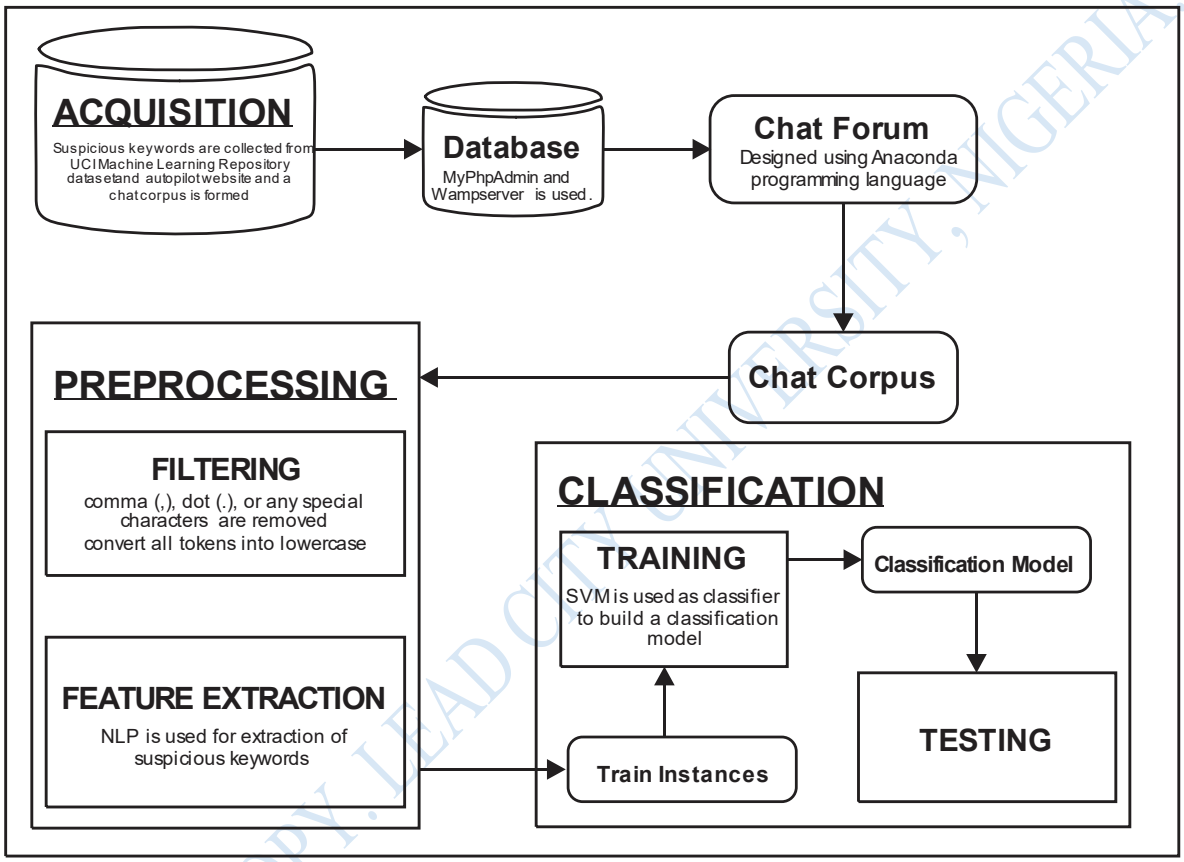

Figure 3.4: Diagram of The Model

3.4 Data Collection Technique

The following sources were used in this project work to collect data. The UCI Machine Learning Repository dataset was used. This is a very good collection of datasets for experimental research purposes. Over 100 spam Trigger words were collected from autopilot website (https://journeys.autopilotapp.com/blog/email-spam-trigger-words/) and

http://archive.ics.uci.edu/ml/datasets/SMS+Spam+Collection from the 202 data spam trigger words available on their database. Discussions took place before and during the study period without the user being given prior knowledge of the study. As a result, the chat is natural and has not been intentionally changed. The chat is written in English format, but the words come from the language of a particular region.

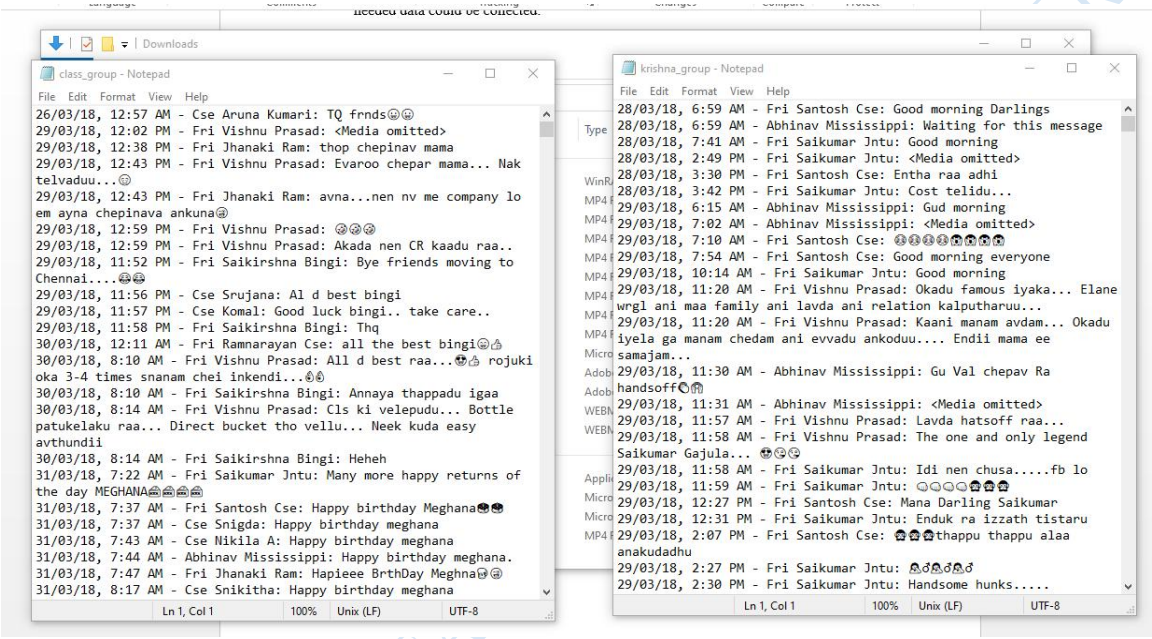

Figure 3.5: Raw Data Collected from UCI Machine Learning Repository

Sources: UCI Machine Learning Repository (2021)

Endnotes

1. A., Rathore, Quantitative Research-Characteristics, 2019.

2. R.M., Miller, C.D. Chan, & L.B., Farmer, Interpretative phenomenological analysis: A contemporary qualitative approach. *Counselor Education and Supervision*, *57*(4), 2018, pp.240-254.

3. J.W. Creswell, & M., Hirose, Mixed methods and survey research in family medicine and community health. *Family Medicine and Community Health*, *7*(2), 2019

Chapter Four

Testing and Evaluation

The foundation of this study has been laid and all necessary explanations and definitions have been provided in previous chapters, Chapter two presented related works from which the insight of this study was developed. Types of data and the various methods of data collection, the various parameters used as input to generate output, the mathematical models used to obtain a remarkable result were clearly presented in chapter three.

This chapter discusses the results of the objectives of which were:

i. to design a model with a chatbox that will enable an admin to monitor suspicious discussions online.
ii. to implement the technique designed in (1) above
iii. to evaluate the performance of the implemented system

It summarily demonstrated the application's menu and their functions and finally concluded with a conclusion.

4.1 Implementation

The project is built using socket programming, in which the client sends a request to the server, and the server responds. StartServer and Thread are the two classes that implement the server. The StartServer command starts a server that listens on a particular port. When a new client is connected, a Thread is generated to serve that client.

Because each connection is handled in its own thread, the server may manage several clients at once. The Thread class is responsible for receiving the messages supplied by the client and broadcasting them to other clients.

Client, Read Thread, and Write Thread are the three classes that make up the client. The client launches the client software and establishes a connection with the server. IP address and port number are used to specify the client's connection. The Read Thread and Write Thread are formed as soon as the connection is established. Until the client is disconnected, the Read Thread reads the input from the server and prints it on the console. The Write Thread receives user input and transmits it to the server.

Analyzesof the characteristics or behavior of each person in the chat or discussion forum, such as their goal and plan when conversing can be done using this chat application. This system includes a server that serves as both an administrator and a client or user. Clients, also known as users, converse with one another, while the server monitors their conversations. Users can talk privately if they don't want their message to be seen to all other users online, or publicly if they don't have any private messages. This chat program aids in the reduction of terrorist actions by monitoring all talks in the discussion forum. This technology has the benefit of monitoring the entire chat without the user's awareness.

If the user or clients use questionable terms, the server will receive an alert message. The server's admin may then monitor or examine the message to see whether the user has used any suspicious phrases. If a user uses questionable terms, that person will be removed from

the system and will not be allowed to login again. If they wish to participate in the chat forum, they must register, sign up, or create a new account.

This system will be able to provide security and dependability to the user and other individuals through this procedure, as well as safeguard the stored data, decreasing terrorist actions.

4.1.1 Running The Developed Suspicious Chat Monitoring System

Step 1: For the program to run, Ampps Server Need be to install and activated to run background.

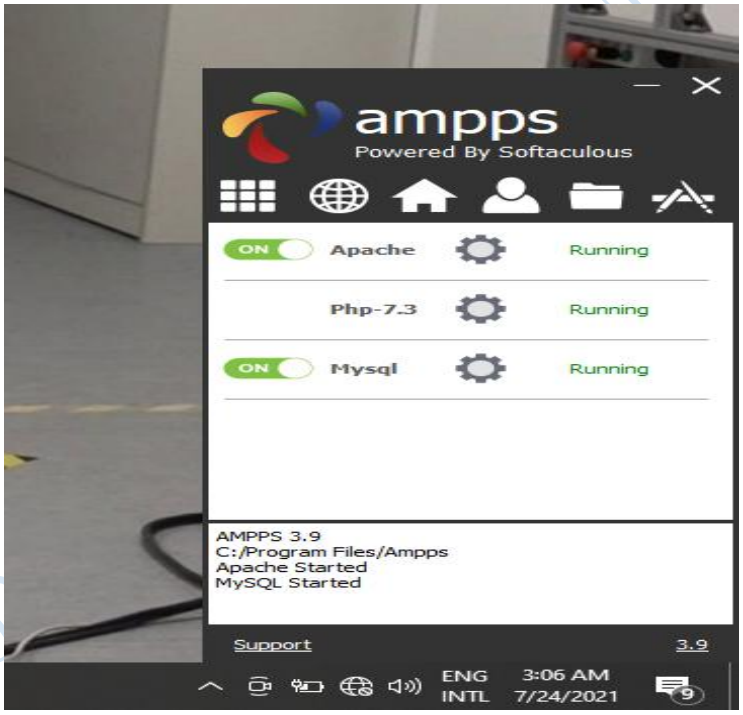

Figure 4.1: Ampps Server Screen

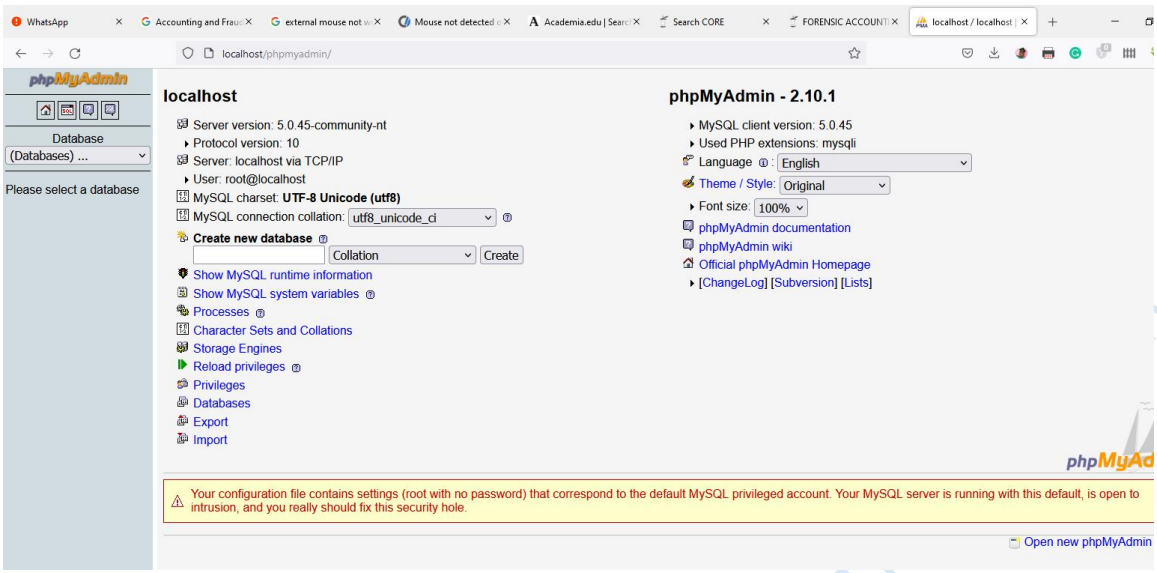

Figure 4.2: Login into MySQL to Create Database

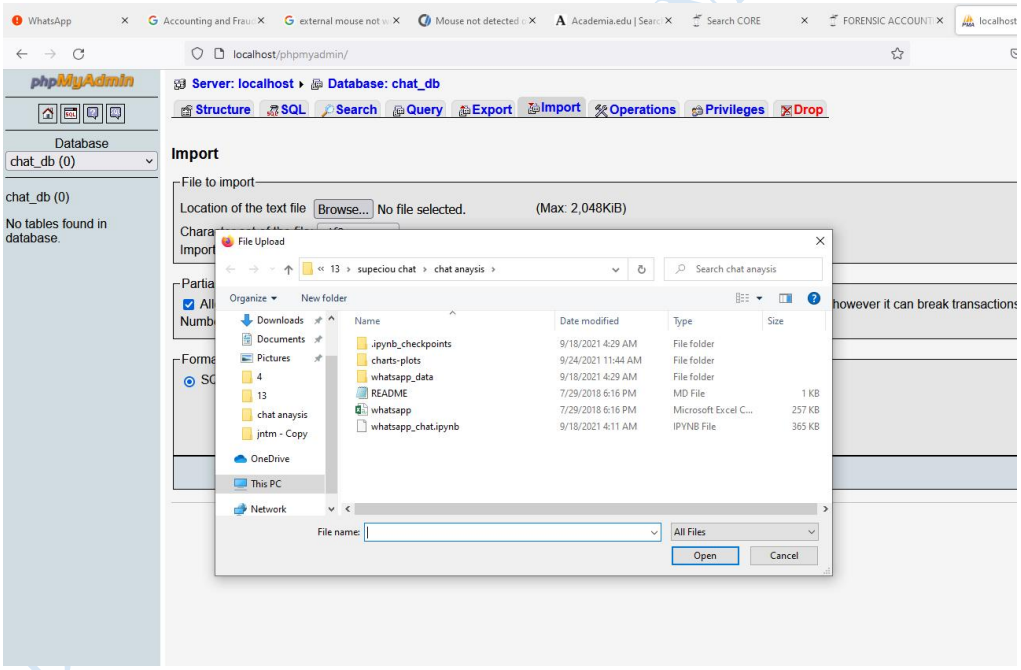

Figure 4.3: Program File is Imported to PhPMyAdmin Local Server

Step 2: Database with Suspicious Keywords Added

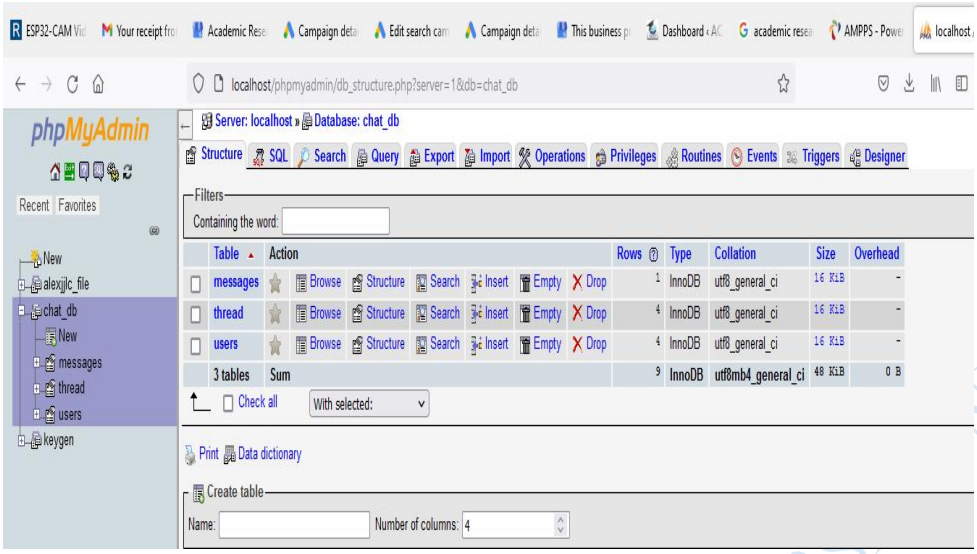

Figure 4.4: Database Created Screen View

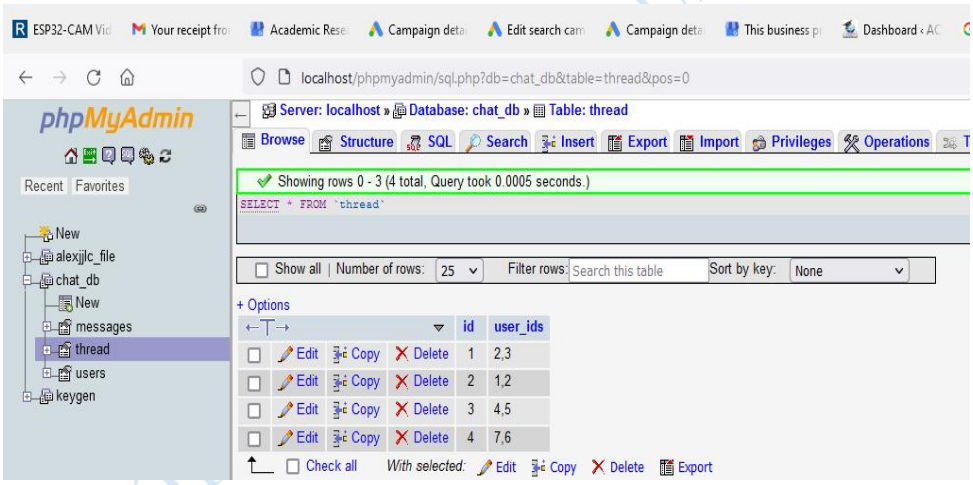

Figure 4.5: Database Editable Option

For Program to Detect and Monitor Suspicious Chat Two More User Need to Create Account with Different Bowers Computer with The Application.

Step 3: Sign Up Screen View

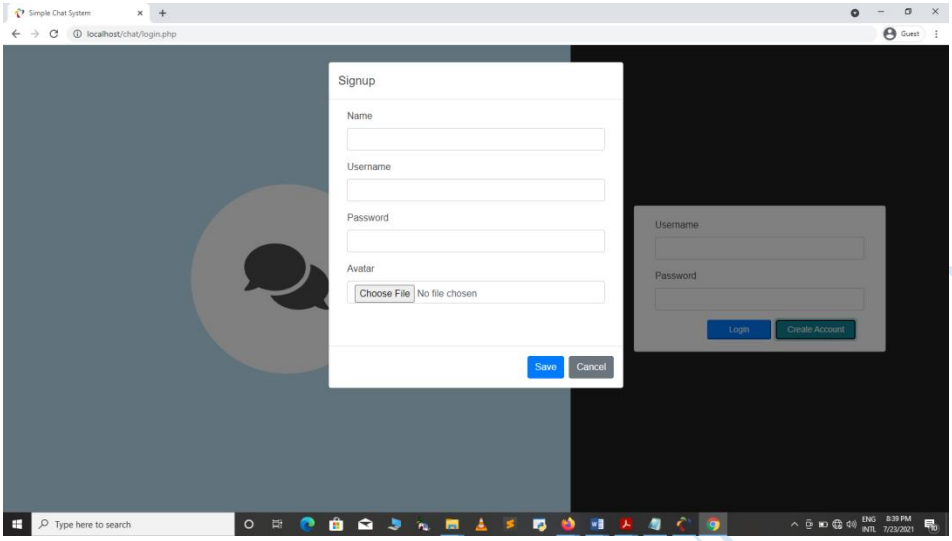

Figure 4.6: User Sign up Screen

After Sign Up User Most Logging with Their Password and User Name Save To Database.

Step 4. User Login Panel

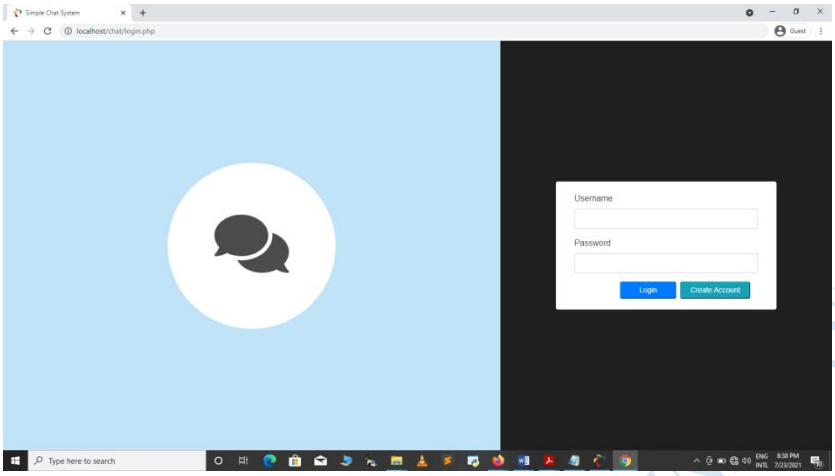

Figure 4.7: User Login Screen

No Number of Active User Online On the Application

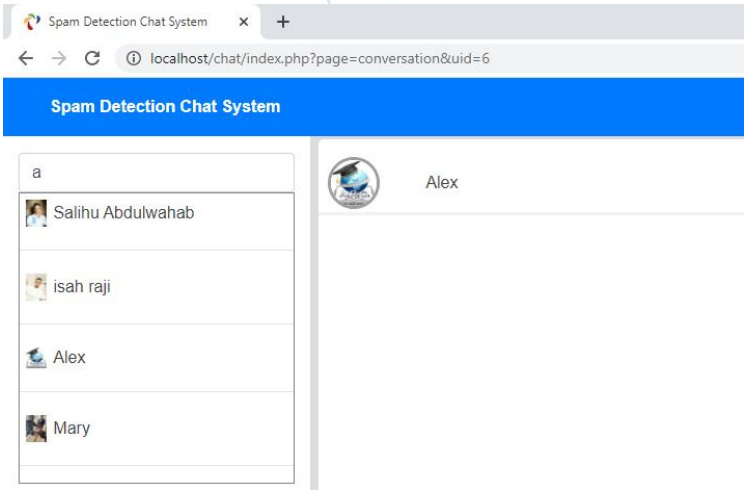

Figure 4.8: Spam Detection Chatting Screen

Step 5: User Chatting with Friend

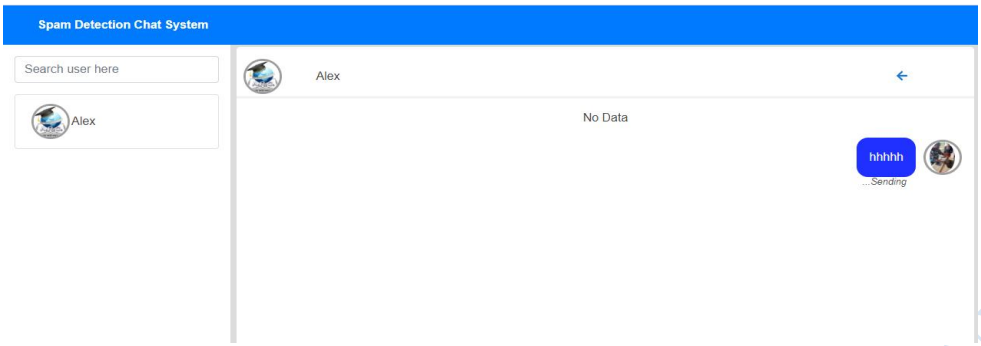

Figure 4.9: User Chatting with Friend

Suspicious Discussions Detected Screen

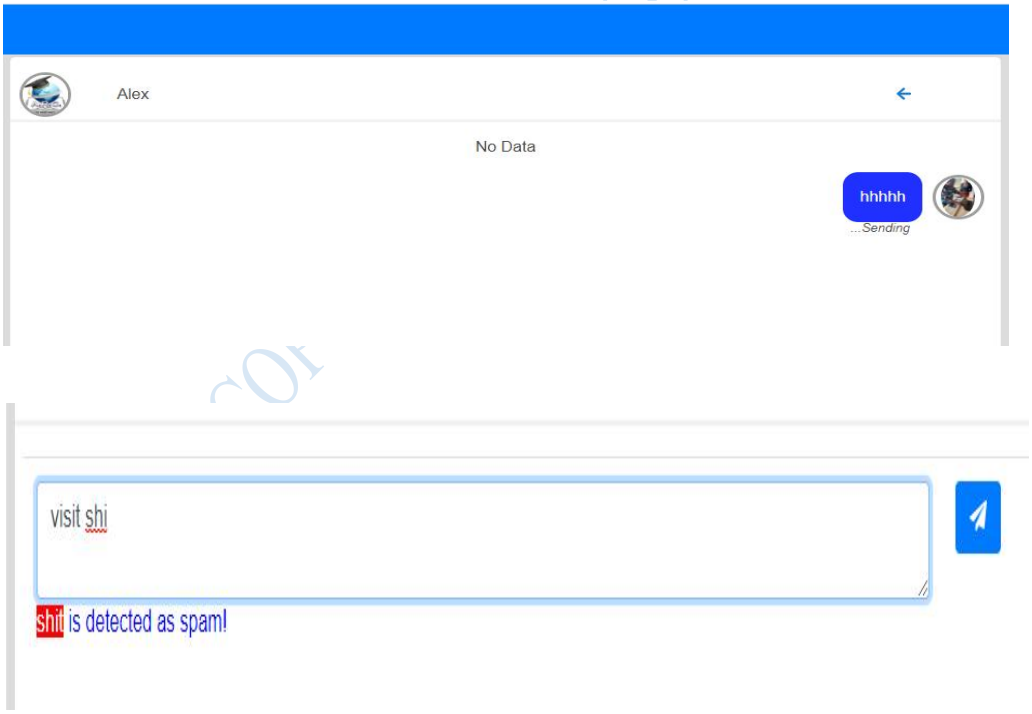

Figure 4.10: Suspicious Discussions Detected Screen

4.2 Discussion of Results

The main aim of this research is to design a technique for monitoring suspicious discussions online forums. This application collects the postings and comments from the discussion sites and analyses those comments using machine learning techniques and algorithm.

Registration

Before login into the chat application, both the admin and the client must first register. Registration is required since without it, the user will not be able to use the chat program.

Login

After completing the registration process, both the server administrator and the client can log in using their registered email addresses and passwords. Admin can monitor chat after login and customer or user can join chat group or chat room and send messages to other customers.

Admin

The administrator's main responsibility is to enter suspicious keywords into the system and monitor customer chats for illegal or suspicious online behavior.

User

The user, also known as the client, is in charge of talking and sending messages via the internet. Depending on the substance of their communications, they can talk privately or openly. These customers are also accountable for employing questionable phrases in online chats.

Keywords

Keywords are a list of suspicious terms supplied by the admin for the purpose of talking. These keywords will be saved in the database, and if the user uses them, an alert message will be sent to the administrator.

Delete

Delete is also controlled by the administrator. If the account is determined to be involved in unlawful behavior, it will be terminated by the administrator. Once a user has been deleted, they must re-register to use the chat system or join the discussion forum.

4.3. Testing and Performance Evaluation

Testing is a technique for finding errors in source code by running it. In order to implement a system, it is necessary to test it first. A software should be thoroughly tested to identify any flaws. Testing is done in tandem with the development of the program, and the more the application is developed, the more testing is required. We can decrease the likelihood of an error or malfunction in the software by testing it. Many parameters and function implementations may change, causing the system to fail to perform as it should, even if the code compiles without problems[1].

We should additionally test the system to ensure that no functions were overlooked throughout the project's execution. Testing takes place once the project is completed. There are many different sorts of tests to guarantee that the system is completely functional and

performing as intended. We'll concentrate on unit testing and integration testing in this section.

Unit testing involves testing an individual component or unit. It is done as part of unit testing phase of software development life cycle and can be done in two phases. We checked the correct names of the applications in the network, then the two-way connection between the server and the client. We can test top-down or bottom-up and then isolate the results through unit testing.

After the unit tests, we did the integration tests. The basic purpose of integration testing is to see if the modules can work well together, that is, to test the interfaces between them. We need to do a full test once the links between the modules are formed. Once the application is complete, system tests will be performed.

4.3.1 Evaluation Using System Usability Scale (SUS)

The new system was implemented with different group of users and was subjected to System Usability Scale (SUS) to determine its level of satisfaction, effectiveness and efficiency. The SUS consists of ten (10) standardized question based on Likert Scale where Strongly Disagree = 1, Disagree = 2, Agree = 3, Strongly Agree = 4.

SUS uses complex scoring system because it comprises of five (5) positive odd numbered questions and five (5) even negative numbered questions.

SUS score = (X + Y) * 2.5 where

X = Add up the total score of all odd numbered questions then subtract 5 while

Y = Add up the total score of all even numbered questions then subtract from 25.

4.3.2 Usability Score for User's Experience

Users	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	SUS Score	NPS
1	4	1	4	1	3	1	4	1	4	2	82.5	Passive
2	4	1	4	1	4	1	4	1	4	1	87.5	Promoter
3	3	1	4	2	4	1	4	1	4	1	82.5	Promoter
4	4	1	3	1	4	2	4	2	4	1	80.0	Promoter
5	4	1	4	2	4	1	3	2	3	1	77.5	Passive
6	4	2	4	1	4	1	4	1	4	1	85.0	Promoter
7	4	2	4	2	4	2	3	2	4	1	75.0	Passive
8	4	1	4	1	4	1	4	1	4	1	87.5	Promoter
9	4	1	4	1	4	1	4	2	4	2	82.5	Promoter
10	4	1	4	2	4	2	4	1	4	1	82.5	Promoter
11	4	2	4	2	3	2	3	2	4	1	72.5	Passive
12	3	2	3	2	4	2	4	2	4	1	72.5	Passive
13	4	1	3	1	4	1	4	2	4	1	82.5	Passive
14	4	1	4	1	4	1	4	1	4	1	87.5	Promoter
15	4	1	3	2	4	2	3	2	3	1	72.5	Passive
16	3	1	4	2	4	1	4	1	4	1	82.5	Promoter
17	4	2	4	2	3	2	3	2	4	1	72.5	Passive
18	4	2	4	2	3	2	3	2	4	1	72.5	Passive
19	4	2	3	2	4	2	4	1	3	1	75.0	Passive
20	4	1	4	1	4	1	4	1	4	1	87.5	Promoter

Table 4.1: SUS Score for User's Experience

Mean SUS Score = $\frac{\text{Sum of all SUS Scores}}{\text{Number of Users}}$

Sum of all SUS Scores for all Users = 82.5 + 87.5 + 82.5 + 80.0 + 77.5 + 85.0 + 75.0 + 87.5 + 82.5 + 82.5 + 72.5 + 72.5 + 82.5 + 87.5 + 72.5 + 82.5 + 72.5 + 72.5 + 75.0 + 87.5

$= \frac{1600}{20}$

The mean SUS Score = 80.0

4.3.3 Interpretation of Result

System Usability Scale (SUS) scores becomes meaningful by normalizing scores to produce percentile ranking. The mean SUS score (80.0) for the system was normalized into percentile ranking of 86%

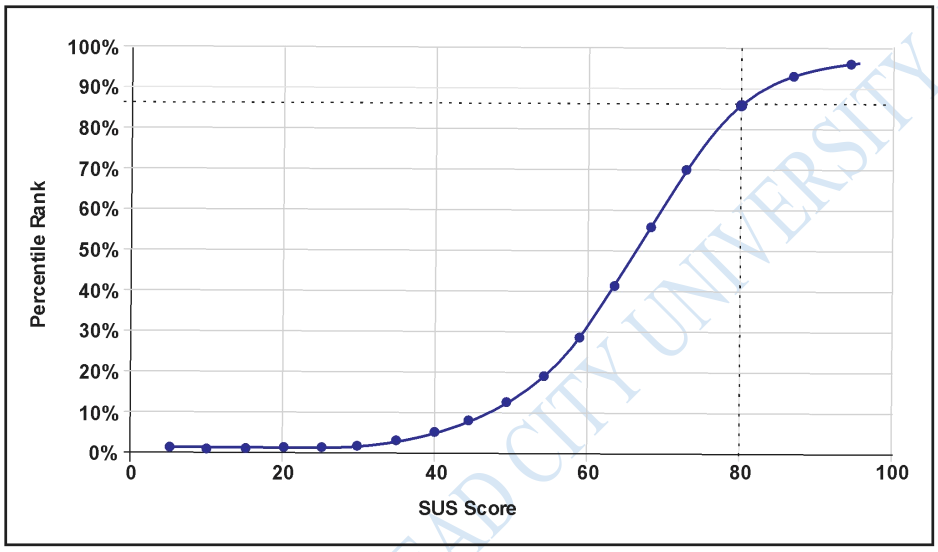

Figure 4.11: Percentile Ranking for Common SUS Scores.

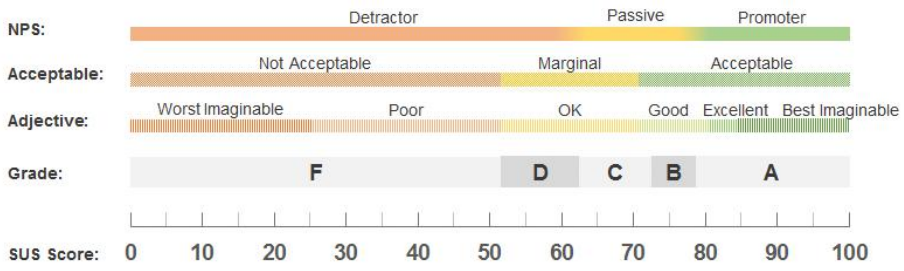

Figure 4.12: Percentiles, Grades, Adjectives, and NPS Categories to Describe Raw SUS Scores.

Source: Online (2021)

4.3.4 Comparative Evaluation Analysis

A comparative Analysis of the system has been carried out to establish the functionality of the study as described in table 4.3 while the bar chart in figure 4.13 Shows the graphical analysis of the system. The study was compared with the Initial System and the result shows that this monitoring suspicious discussion system has an edge over the existing system.

	Chat Monitoring System	Spam Detection System	Suspicious Keywords Detection System
SUS Scores	80.0	76.2	71.2
2	87.5	70.5	70.5
3	82.5	72.5	72.5
4	80.0	70.5	65.5
5	77.5	70.0	70.0
6	85.0	80.0	80.0
7	75.0	65.5	70.5
8	87.5	70.5	65.5
9	82.5	75.9	75.9
10	82.5	85.5	85.5
11	72.5	80.5	80.5
12	72.5	80.5	80.5
13	82.5	80.0	70.5
14	87.5	75.9	75.9
15	72.5	60.0	80.0
16	82.5	60.5	60.5
17	72.5	70.0	60.0
18	72.5	70.0	70.5
19	75.0	70.0	70.0
20	87.5	80.5	60.5

Table 4.2 SUS Scores for Comparative Evaluation

From the result in table 4.3 The raw and mean SUS score is 80.9 for the cybercrime monitoring system. It was normalized to percentile ranking 90. This indicate that the system was excellent and acceptable and the users were promoters. The users will not discourage others from using the proposed system while the mean SUS score for the initial system is 75.25 was normalized to percentile ranking 72. This indicate that the system is acceptable and the users were passive. The users will discourage others from using the system.

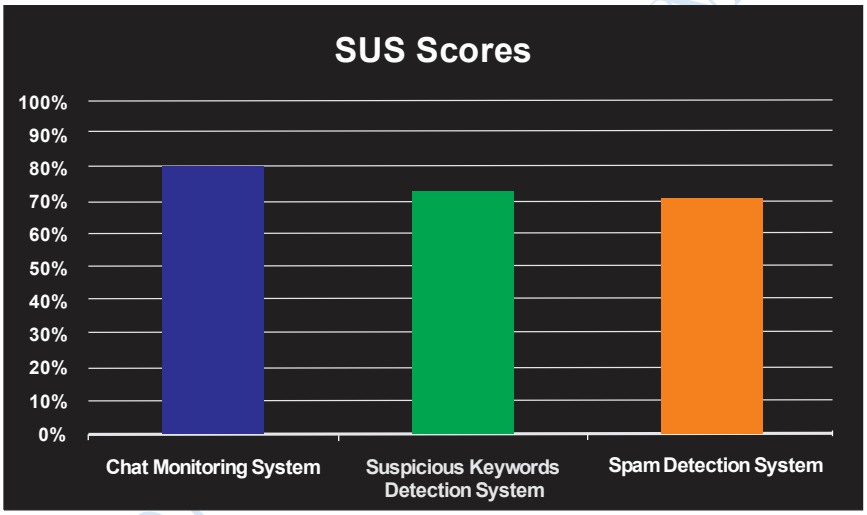

Figure 4.13: SUS Scores Bar Chat for Comparative Evaluation

Endnote

1. Y. Zhao, Z. Wang, J. Chen, M. Liu, M. Wu, Y. Zhang & L. Zhang, History-Driven Test Program Synthesis for JVM Testing. *IEEE/ACM 44th International Conference on Software Engineering (ICSE)* (pp. 1133-1144). 2022, IEEE.

Chapter Five

Conclusion

5.1. Summary of Findings

Human behavior and social relationships are being influenced by digital and multimedia technologies. This technology will be used to create an automated system for identifying suspicious talks on internet forums, allowing us to expose all users' questionable actions and interests. Due to a lack of information retrieval analysis from online forums, the true efficiency of online forums retrieval data analysis remains disputed. The goal of this technology is to automatically monitor questionable comments on internet forums. Suspicious posts on internet forums are detected using text analysis. This system also prioritizes plan execution time and automatic categorization to detect more significant suspicious conversations.

However, as the internet becomes more and more popular, various cybercrimes such as fraud, unauthorized posting and other illegal activities are on the rise. This article presents a system that not only detects and reports illegal activity on online discussion forums, but also helps mitigate it by restricting the content that users can post publicly. Several text mining algorithms are used in our system to filter out illegal and fraudulent messages, resulting in a legitimate platform for users to express themselves.

5.2 Conclusion

Suspect online discussions will be monitored using deep learning techniques, which will evaluate plain text and recognize suspicious phrases. Users can speak with one another in this

system, and the administrator keeps an eye on the conversation. This method may also be used to detect suspicious phrases spoken between users and prevent them from being used again by removing them from the server database. Because the target users would be unaware that they are being tracked, it will be easier to maintain track of the suspects without their noticing and have complete control over them. This method will minimize unlawful actions and, more importantly, the number of users who become involved in such activities. As a result, this system guarantees protection for users while also allowing administrators to see what plans are being made among the users.

The overall process of active conversation monitoring and suspicious chat detection over the internet has been completed, and it has shown to be beneficial to individuals. The aforementioned analyses and works provide proof of this. If the future upgrades are done correctly, the research work can be extended in the future. The research, titled Monitoring Suspicious Online Discussions Using Deep Learning Techniques, has been evaluated using sample data and proven to be effective.

The system was created with the users/people in mind. The database approach to system development has helped to reduce data redundancy and improve data consistency in the system. This method is adaptable and simple to use. The system meets all of the client's needs.

5.3 Recommendations

i. The system be incorporated with Microsoft Corporation so that it comes by default with Windows of computer system.

ii. More research and more effort on development of the program so that the Forensic Interface does not pop up on running the Projects on Netbeans rather remain hidden.

iii. With more development of the project, the project can also help in capturing the letters typed by the user while on the computer system and also take snapshot of the user at a particular time interval.

5.4 Contribution to Knowledge

Having discovered the limitations of the existing systems, this research improves the performance and the security over the existing systems.

1. This research will contribute theoretically to the existing body of knowledge through a more extensive understanding of emerging technologies, with potential for future research.
2. It will also contribute essentially and experientially through a developed and dependable robust model.
3. This chat program aids in the reduction of terrorist actions by monitoring all talk in the discussion forum. This technology has the benefit of monitoring the entire chat without the user's awareness.
4. This system can make society more stable by reducing the number of crimes, and it will also provide security and protection for users.

5.5 Suggested Area for Further Research

In the future, we plan to train the model on larger datasets to improve overall performance. Suspicious text subdomains are considered to make the dataset more diverse. Furthermore,

recurrent learning algorithms can be employed to capture the inherent sequential patterns of long texts.

Bibliography

Books

Brindha, M., Vishnupriya, V., Rohini, S., Udhayamoorthi, M. & Mohan, K.S., Active Chat Monitoring and Suspicious Detection Over Internet, 2018.

Abdul-Rahim, M.H., Assessment of the Perceptions on the Effects of Cybercrime on Senior High Students in Tamale Metropolis: A Case Study of Vitting Senior High School *(Doctoral Dissertation)*, 2021.

Esteves, R.M.F., Financial Digitalization: Risk or Opportunity *(Doctoral Dissertation),* 2021.

Holt, T. & Bossler, A., Cybercrime in Progress: Theory and Prevention of Technology-Enabled Offenses. *Routledge.* 2015.

Hoogeveen, D., Real and Misflagged Duplicate Question Detection in Community Question-Answering *(Doctoral Dissertation),* 2018.

Ugbe, U.M., Exploring The Security Measures to Reduce Cyberattacks Within the Nigerian Banking Sector: A Qualitative Inquiry *(Doctoral Dissertation, Capella University),* 2021.

Journals

Aggarwal, K., Mijwil, M.M., Al-Mistarehi, A.H., Alomari, S., Gök, M., Alaabdin, A.M.Z. & Abdulrhman, S.H., Has The Future Started? The Current Growth of Artificial Intelligence, Machine Learning, And Deep Learning. *Iraqi Journal for Computer Science and Mathematics,* 3(1), 2022, pp.115-123.

Ahmad, B., Jian, W. & Anwar Ali, Z., Role of Machine Learning and Data Mining in Internet Security: Standing State with Future Directions. *Journal of Computer Networks and Communications,* 2018.

Almatrafi, O. & Johri, A., Systematic Review of Discussion Forums in Massive Open Online Courses (MOOCs). *IEEE Transactions on Learning Technologies,* 12(3), 2018, pp.413-428.

Almeida, L.Á., Data Model Classification Based On Machine Learning Techniques for Detection of Anomalous Traffic. *Cartagena de Indias,* 2019.

Avaghade, S.B., Rajkumar, A.V., Jayvant, P.J., Prithviraj, S.S., Dilip, P.A. & Dacoe, K., Active Chat Monitoring and Suspicious Detection Over Internet. *International*

Journal of Engineering Applied Sciences and Technology, Vol. 4, Issue 11, ISSN No. 2455-2143, 2018, pages 397-399

Batra, S., Gupta, M., Singh, J., Srivastava, D. & Aggarwal, I., An Empirical Study of Cybercrime and Its Preventions. *Sixth International Conference On Parallel, Distributed and Grid Computing (PDGC),* 2020, (pp. 42-46). IEEE.

Berk, R.A., Artificial Intelligence, Predictive Policing, And Risk Assessment for Law Enforcement. *Annual Review of Criminology,* 4, 2021, pp.209-237.

Beutel, I., Kirschler, O. & Kokott, S., How Do Fake News and Hate Speech Affect Political Discussion and Target Persons and How Can They Be Detected? *Central and Eastern European eDem and eGov Days,* 342, 2022, pp.37-81.

Bhamare, B.R. & Prabhu, J., A Supervised Scheme for Aspect Extraction in Sentiment Analysis Using the Hybrid Feature Set of Word Dependency Relations and Lemmas. *PeerJ Computer Science,* 7, 2021, p.e347.

Briciu, V.A. & Briciu, A., Social Media and Organizational Communication. *In Encyclopedia of Organizational Knowledge, Administration, And Technology,* 2021 (pp. 2609-2624). IGI Global.

Canter, L., The Misconception of Online Comment Threads: Content and Control On Local Newspaper Websites. *Journalism Practice,* 7(5), 2013, pp.604-619.

Caramancion, K.M., Li, Y., Dubois, E. & Jung, E.S., The Missing Case of Disinformation from The Cybersecurity Risk Continuum: A Comparative Assessment of Disinformation with Other Cyber Threats. *Data,* 7(4), 2022, p.49.

Chen, J. & Neo, P., Texting The Waters: An Assessment of Focus Groups Conducted Via the Whatsapp Smartphone Messaging Application. *Methodological Innovations,* 12(3), 2019, p.2059799119884276.

Chen, R.C., Dewi, C., Huang, S.W. & Caraka, R.E., Selecting Critical Features for Data Classification Based On Machine Learning Methods. *Journal of Big Data,* 7(1), 2020, pp.1-26.

Creelman, V., "Thank You for Reaching Out:" Brand Relationship Management and The Conversational Human Voice of Customer Care in Online Service Encounters. *Discourse, Context & Media,* 46, 2022, p.100572.

Creswell, J.W. & Hirose, M., Mixed methods and survey research in family medicine and community health. Family Medicine and Community Health, *7*(2), 2019

Crisci, C., Ghattas, B. & Perera, G., A Review of Supervised Machine Learning Algorithms and Their Applications to Ecological Data. *Ecological Modelling,* 240, 2012, pp.113-122.

Dam, T., Klausner, L.D., Buhov, D. & Schrittwieser, S., Large-Scale Analysis of Pop-Up Scam On Typosquatting Urls. *In Proceedings of The 14th International Conference On Availability, Reliability and Security,* 2019, (pp. 1-9).

Dang, Q.V., Studying Machine Learning Techniques for Intrusion Detection Systems. *In International Conference On Future Data and Security Engineering,* 2019, (pp. 411-426). Springer, Cham.

Daugirdas, K., Reputation as A Disciplinarian of International Organizations. *American Journal of International Law,* 113(2), 2019, pp.221-271.

Davis, T.C., Wolf, M.S., Bass III, P.F., Thompson, J.A., Tilson, H.H., Neuberger, M. & Parker, R.M., Literacy and Misunderstanding Prescription Drug Labels. *Annals of Internal Medicine,* 145 (12), 2006, pp.887-894.

Dhanabal, L. & Shantharajah, S.P., A Study On NSL-KDD Dataset for Intrusion Detection System Based On Classification Algorithms. *International Journal of Advanced Research in Computer and Communication Engineering,* 4(6), 2015, pp.446-452.

Dutta, N., Jadav, N., Tanwar, S., Sarma, H.K.D. & Pricop, E., Intrusion Detection Systems Fundamentals. *In Cyber Security: Issues and Current Trends,* 2022, (pp. 101-127). Springer, Singapore.

Enke, N. & Borchers, N.S., Social Media Influencers in Strategic Communication: A Conceptual Framework for Strategic Social Media Influencer Communication. *In Social Media Influencers in Strategic Communication,* 2021 (pp. 7-23). Routledge.

Gandhi, F., Pansaniya, D. & Naik, S., Ethical Hacking: Types of Hackers, Cyber Attacks and Security. *International Research Journal of Innovations in Engineering and Technology,* 6(1), 2022, p.28.

Ghafory, H., "Monitoring Suspicious Communication On Online." *International Journal of Advanced Academic Studies;* 2(2): 66-69, 2020.

Habib, A., Hossain, F., Ferdous, T. & Bayezid, K.M., Social Networks and Social Ties: Changing Trends Among Urban Dwellers in Bangladesh. *Open Access Library Journal,* 5(5), 2018, pp.1-12.

Hanley, T., Prescott, J. & Gomez, K.U., A Systematic Review Exploring How Young People Use Online Forums for Support Around Mental Health Issues. *Journal of Mental Health,* 28(5), 2019, pp.566-576.

Hasan, A.M., Rassem, T.H., Noor, N.M. & Hasan, A.M., A Semantic Taxonomy for Weighting Assumptions to Reduce Feature Selection from Social Media and Forum Posts. *In International Conference of Reliable Information and Communication Technology,* 2019, (pp. 407-419). Springer, Cham.

Hoque, M.A. & Davidson, C., Design and Implementation of an IoT-Based Smart Home Security System. *Int. J. Networked Distributed Comput.,* 7(2), 2019, pp.85-92.

Houssein, E.H., Kilany, M. & Hassanien, A.E., ECG Signals Classification: A Review. *International Journal of Intelligent Engineering Informatics,* 5(4), 2017, pp.376-396.

Howell, C.J., Burruss, G.W., Maimon, D. & Sahani, S., Website Defacement and Routine Activities: Considering The Importance of Hackers' Valuations of Potential Targets. *Journal of Crime and Justice,* 42(5), 2019, pp.536-550.

Ishii, K., Lyons, M.M. & Carr, S.A., Revisiting Media Richness Theory for Today and Future. *Human Behavior and Emerging Technologies,* 1(2), 2019, pp.124-131.

Jia, Y., Zhong, F., Alrawais, A., Gong, B. & Cheng, X., Flowguard: An Intelligent Edge Defense Mechanism Against IoT DDoS Attacks. *IEEE Internet of Things Journal,* 7(10), 2020, pp.9552-9562.

Koi-Akrofi, G.Y., Koi-Akrofi, J., Odai, D.A. & Twum, E.O., Global Telecommunications Fraud Trend Analysis. *International Journal of Innovation and Applied Studies,* 25(3), 2019, pp.940-947.

Kumar, V., Das, A.K. & Sinha, D., Statistical Analysis of The UNSW-NB15 Dataset for Intrusion Detection. *In Computational Intelligence in Pattern Recognition,* 2020, (pp. 279-294). Springer, Singapore.

Kumari, A. & Balkishan, Detection of Suspicious Text Messages and Profiles Using Ant Colony Decision Tree Approach. *International Journal of Business Intelligence and Data Mining,* 19(4), 2021, pp.418-442.

Kwon, N., Deshpande, H., Hasan, M.K., Darnal, A. & Kim, J., Multi-ttach: Techniques to Enhance Multi-Material Attachments in Low-cost FDM 3D Printing. *In Symposium on Computational Fabrication,* 2021, (pp. 1-16).

Lee, H. & Choi, K.S., Interrelationship Between Bitcoin, Ransomware, And Terrorist Activities: Criminal Opportunity Assessment Via Cyber-Routine Activities Theoretical Framework. *Victims & Offenders,* 16(3), 2021, pp.363-384.

Leithner, M. & Simos, D.E., *CHIEv:* Concurrent Hybrid Analysis for Crawling and Modeling of Web Applications. *ACM SIGAPP Applied Computing Review,* 21(1), 2021, pp.5-23.

Li, J., Xu, Q., Cuomo, R., Purushothaman, V. & Mackey, T., Data Mining and Content Analysis of the Chinese Social Media Platform Weibo During the Early COVID-19 Outbreak: Retrospective Observational Infoveillance Study. *JMIR Public Health and Surveillance,* 6(2), 2020, p.e18700.

Lin, H., Yan, Z., Chen, Y. & Zhang, L., A Survey On Network Security-Related Data Collection Technologies. *IEEE Access,* 6, 2018, pp.18345-18365.

Messerschmidt, J.W., From Marx to Bonger: Socialist Writings On Women, Gender, And Crime. *Sociological Inquiry,* 58(4), 1988, pp.378-392.

Miller, R.M., Chan, C.D. & Farmer, L.B., Interpretative phenomenological analysis: A contemporary qualitative approach. *Counselor Education and Supervision*, *57*(4), 2018, pp.240-254.

Ming, Y., Qu, H. & Bertini, E., Rulematrix: Visualizing and Understanding Classifiers with Rules. *IEEE Transactions On Visualization and Computer Graphics,* 25(1), 2018, pp.342-352.

Morris, J.W., Hearing The Past: The Sonic Web from MIDI to Music Streaming". The SAGE Handbook of Web History. Londres: *SAGE Publications,* 2019, pp.491-504.

Nguyen, G., Dlugolinsky, S., Bobák, M., Tran, V., Lopez Garcia, A., Heredia, I., Malík, P. & Hluchý, L., Machine Learning and Deep Learning Frameworks and Libraries for Large-Scale Data Mining: A Survey. *Artificial Intelligence Review,* 52(1), 2019, pp.77-124.

Nobari, A.D., Sarraf, M.H.K.M., Neshati, M. & Daneshvar, F.E., Characteristics of Viral Messages On Telegram; The World's Largest Hybrid Public and Private Messenger. *Expert Systems with Applications,* 168, 20220 p.114303.

Obeidat, I., Hamadneh, N., Alkasassbeh, M., Almseidin, M. & AlZubi, M., Intensive Pre-Processing of KDD Cup 99 For Network Intrusion Classification Using Machine Learning Techniques, 2019.

Ojugo, A.A. & Yoro, R.E., Forging A Deep Learning Neural Network Intrusion Detection Framework to Curb the Distributed Denial of Service Attack. *International Journal of Electrical and Computer Engineering,* 11(2), 2021, p.1498.

Osang, F.B. & Nwaocha, V., Bridging The Distance in Open and Distance Learning: Developing Student-Student and Student-Lecturer Collaborative Forum. *Lautech Journal of Engineering and Technology,* 12(1), 2018, pp.97-112.

Pandey, S., Singh, R.K., Gunasekaran, A. & Kaushik, A., Cyber Security Risks in Globalized Supply Chains: Conceptual Framework. *Journal of Global Operations and Strategic Sourcing,* 2020.

Patel, H.H. & Prajapati, P., Study and Analysis of Decision Tree Based Classification Algorithms. *International Journal of Computer Sciences and Engineering,* 6(10), 2018, pp.74-78.

Paul, P. & Aithal, P.S., Cybercrime: Challenges, Issues, Recommendation and Suggestion in Indian Context. *International Journal of Advanced Trends in Engineering and Technology. (IJATET),* 3(1), 2018. pp.59-62.

Rathore, A., Quantitative Research-Characteristics, 2019.

Rawat, R., Mahor, V., Chirgaiya, S. & Rathore, A.S., Applications of Social Network Analysis to Managing the Investigation of Suspicious Activities in Social Media Platforms. *In Advances in Cybersecurity Management,* 2021 (pp. 315-335). Springer, Cham.

Risch, J. & Krestel, R., Toxic Comment Detection in Online Discussions. *In Deep Learning-Based Approaches for Sentiment Analysis,* 2020, (pp. 85-109). Springer, Singapore.

Roelen-Blasberg, T., Habel, J. & Klarmann, M., Automated Inference of Product Attributes and Their Importance from User-Generated Content: Can We Replace Traditional Market Research? *International Journal of Research in Marketing,* 2022.

Saba, T., Sadad, T., Rehman, A., Mehmood, Z. & Javaid, Q., Intrusion Detection System Through Advance Machine Learning for The Internet of Things Networks. *IT Professional,* 23(2), 2021, pp.58-64.

Sahani, R., Rout, C., Chandrakanta Badajena, J., Jena, A.K. & Das, H., Classification of Intrusion Detection Using Data Mining Techniques. *In Progress in Computing, Analytics and Networking,* 2018, (pp. 753-764). Springer, Singapore.

Schlackl, F., Link, N. & Hoehle, H., Antecedents and Consequences of Data Breaches: A Systematic Review. *Information & Management,* 2022, p.103638.

Shah, A. & Chudasama, D., Investigating Various Approaches and Ways to Detect Cybercrime. *Journal of Network Security,* 9(2), 2021, pp.12-20p.

Slavko, A.S., Zavhorodnia, V.M. & Shevchenko, N.Y.A., Protection of One's Honor, Dignity, and Business Reputation on Social Networks: Issues and Ways to Resolve Them, 2020.

Srivastava, T., Mangalagowri, R. & Dudala, S.S. Monitoring of Suspicious Discussions on Online Forums Using Data Mining. *International Journal of Pure and Applied Mathematics,* 118(22), 2018, pp.257-262.

Tariq, H., Hanif, M.K., Sarwar, M.U., Bari, S., Sarfraz, M.S. & Oskouei, R.J. Employing Deep Learning and Time Series Analysis to Tackle the Accuracy and Robustness of the Forecasting Problem. *Security and Communication Networks,* 2021.

Terziev, V., Shared Knowledge and Presentation of the Research Results in Specialised Scientific Publications. *Available at SSRN,* 2022.

Thomas, J.K., Redhya, M. & Mahalekshmi, T., Monitoring Suspicious Discussions in Online Forums Using Data Mining. *International Journal of Research Publication and Reviews,* Vol 2, No 12, 2021, pp 682-688, ISSN 2582.

Thuraisingham, B.M., Can AI be for Good in the Midst of Cyber Attacks and Privacy Violations? A Position Paper. *In Proceedings of the Tenth ACM Conference on Data and Application Security and Privacy,* 2020, (pp. 1-4).

Thylstrup, N.B., The Ethics and Politics of Data Sets in The Age of Machine Learning: Deleting Traces and Encountering Remains. *Media, Culture & Society,* Vol. 44 (4) 855671, 2022, p.01634437211060226.

Wellington, K., Cyberattacks On Medical Devices and Hospital Networks: Legal Gaps and Regulatory Solutions. *Santa Clara High Tech. LJ,* 30, 2013, p.139.

Wong, Y.K. Advanced Deep Learning Approach and Applications. *International Journal of Information Technology (IJIT),* 7(5). 2021.

Xue, A.Y., Qi, J., Xie, X., Zhang, R., Huang, J. & Li, Y., Solving The Data Sparsity Problem in Destination Prediction. *The VLDB Journal,* 24(2), 2015, pp.219-243.

Zhao, N. & Lu, J., Review of Neural Network Algorithm and Its Application in Temperature Control of Distillation Tower. *Journal of Engineering Research and Reports,* 20(4), 2021, pp.50-61.

Zhao, Y., Wang, Z., Chen, J., Liu, M., Wu, M., Zhang, Y. & Zhang, L., History-Driven Test Program Synthesis for JVM Testing. *IEEE/ACM 44th International Conference on Software Engineering (ICSE)* (pp. 1133-1144). 2022, IEEE.

Zhiqiang, L., Mohi-Ud-Din, G., Bing, L., Jianchao, L., Ye, Z. & Zhijun, L., Modeling Network Intrusion Detection System Using Feed-Forward Neural Network Using UNSW-NB15 Dataset. *IEEE 7th International Conference on Smart Energy Grid Engineering (SEGE),* 2019, (pp. 299-303). IEEE.

Zhou, X., Lu, P., Zheng, Z., Tolliver, D. & Keramati, A., Accident Prediction Accuracy Assessment for Highway-Rail Grade Crossings Using Random Forest Algorithm Compared with Decision Tree. *Reliability Engineering & System Safety, 200,* 2020, p.106931.

Appendices

Appendix I

Admin Code

```
<?php
		$data = " name = '$name' ";
		$data .= ", username = '$username' ";
		if(!empty($password))
		$data .= ", password = '".md5($password)."' ";
		$chk = $this->db->query("Select * from users where username =
'$username' and
		}
		if($save){
			return 1;
		}
	}
	function create_account(){
		extract($_POST);
		$data = " name = '$name' ";
		$data .= ", username = '$username' ";
		$data .= ", password = '".md5($password)."' ";
		$chk = $this->db->query("SELECT * FROM users where username =
'$username' ")->num_rows;
		if($chk > 0){
			return json_encode(array("status"=>2,"msg"=>"Username already
exist."));
			exit;
		}
		if($_FILES['img']['tmp_name'] != ''){
					$fname = strtotime(date('y-m-d
H:i')).'_'.$_FILES['img']['name'];
					$move =
move_uploaded_file($_FILES['img']['tmp_name'],'assets/uploads/'. $fname);
				$data .= ", avatar = '$fname' ";

		}

	}
	function send_chat(){
		extract($_POST);
		$data = " message = '$message' ";
		$data .= ", user_id = '{$_SESSION['login_id']}' ";
		if(empty($convo_id)){
			$cdata = " user_ids = '$user_id,{$_SESSION['login_id']}' ";
```

```
                        $cdata2 = " user_ids = '{$_SESSION['login_id']},$user_id' ";
                        $user_ids = $_SESSION['login_id'].",".$user_id;
                        $chk = $this->db->query("SELECT * from thread where $cdata or
$cdata2 ");
                        if($chk->num_rows> 0){
                                $convo_id = $chk->fetch_array()['id'];
                        }else{
                                $thread = $this->db->query("INSERT INTO thread set
$cdata ");
                                $convo_id = $this->db->insert_id;
                        }
                }else{
                        $qry =$this->db->query("SELECT * from thread where md5(id)
='$convo_id' ")->fetch_array();
                        $convo_id = $qry['id'];
                        $user_ids = $qry['user_ids'];
                }
                $data .= ", convo_id = '$convo_id' ";
                $save = $this->db->query("INSERT INTO messages set $data");
                if($save)
                        return
json_encode(array('status'=>1,'convo_id'=>md5($convo_id),'convo_users'=>$user_ids));
        }
        function load_convo(){
                extract($_POST);
                $data = array();
                $get = $this->db->query("SELECT m.message,u.id,u.name,u.avatar
FROM messages m inner join users u on u.id = m.user_id where md5(m.convo_id) =
'$convo_id' ");
                while($row= $get->fetch_assoc()){
                        $data[] = $row;
                }
                return json_encode($data);
        }
        function read_msg(){
                extract($_POST);
                if(isset($user_id) && $user_id > 0){
                        $update = $this->db->query("UPDATE messages set status = 1
where md5(convo_id) = '$convo_id' and user_id=$user_id ");
                        if($update){
                                return 1;
                        }
                }
        }
}
```

Appendix II

Control Panel

```
<?php
include 'db_connect.php'
?>
<div class="container-fluid pt-3" id="contact-panel">
	<div class="col-lg-12">
		<div class="row">
			<form action="" id="filter-contact" class=" w-100">
				<div class="form-group" id="filter-row">
					<input type="text" class="form-control"
placeholder="Search user here" id="find_user">
					<div id="filtered-field" style="display: none">

					</div>
				</div>
			</form>
		</div>

	</div>
<ul class="list-group" id="convo-list">
	<?php
	$thread = $conn->query("SELECT * from thread where
concat('[',REPLACE(user_ids,',','],['),']') like '%[{$_SESSION['login_id']}]%' ");
	while($row= $thread->fetch_array()):
		$user = $conn->query("SELECT * FROM users where id in
({$row['user_ids']}) and id!= {$_SESSION['login_id']} ")->fetch_array();
		$msg = $conn->query("SELECT * FROM messages where convo_id =
{$row['id']} and user_id='{$user['id']}' and status = 0 order by id desc limit 1 ")-
>num_rows;
	?>
	<li class="list-group-item w-100" data-id="<?php echo md5($row['id']) ?>">
		<table class="" width="100%">
			<tr>
				<td width="10%"><div class="uavatar"><img
src="assets/uploads/<?php echo $user['avatar'] ?>" alt=""></div></td>
				<td><?php echo ucwords($user['name']) ?><span
class="badge badge-danger float-right notif"><?php echo $msg ?></span></td>
			</tr>
		</table>
	</li>
<?php endwhile; ?>
</ul>
	<div id="search_clone" style="display: none">
```

```
                <div class="searched-user">
                        <table width="100%">
                                <tr>
                                        <td width="10%"><img src="" alt=""></td>
                                        <td width="90%"><span></span></td>
                                </tr>
                        </table>
                        <hr>
                </div>
        </div>
</div>
<ul id="clone_ul" style="display: none">
        <li class="list-group-item w-100" >
                <table class="" width="100%">
                        <tr>
                                <td width="10%"><div class="uavatar"><img src=""
alt=""></div></td>
                                <td><span class="uname"></span><span class="badge
badge-danger float-right notif">0</span></td>
                        </tr>
                </table>
        </li>
</ul>
<style>
.notif {
        position:relative;
        top:-15px;
        display:none;
}
        div#contact-panel {
position: fixed;
background: white;
height: calc(100% - 4rem);
width: 18.5rem;
top: 3.5rem;
left: 0;
}
#convo-list li{
        cursor:pointer
}
#convo-list li:hover{
        cursor:pointer;
        background: #008bff47;
}
#convo-list img{
                max-width: calc(100%);
```

```
                max-height:calc(100%);
                border-radius: 100%;
        }
                #convo-list .uavatar {
        width: 50px;
        height: 50px;
        align-self: center;
        border-radius: 50%;
        border: 3px solid #808080c2;
        display: flex;
        justify-content: center;
        text-align: -webkit-auto;
        }
#fileter-row{
        position: relative;
}
.searched-user img{
max-width: calc(80%);
max-height: calc(20vh);
margin: 5px;
}
.searched-user span{
        padding-left: 5px;
}
.searched-user{
        cursor:pointer;
}
.searched-user:hover{
        background: #008bff47;
        color: white;
}
#filtered-field{
        position: absolute;
        background: white;
        border:1px solid gray;
width: calc(100%);
z-index: 10
}
</style>
<script>
        $(document).ready(function(){
                var li = $("#convo-list")
                li.find('.notif').each(function(){
                        if($(this).html() > 0){
                                $(this).show()
                        }
```

```
                })
        })
        $('#find_user').on('keyup keydown keypress change',function(){
                if($(this).val() == ''){
                        $('#filtered-field').toggle(false)
                }else{
                        $('#filtered-field').toggle(true)

user').clone()

        div.find('img').attr('src','assets/uploads/'+resp[k].avatar)
                                                        div.find('span').html(resp[k].name)
                                                        div.attr('data-id',resp[k].id)
                                                        $('#filtered-field').append(div)
                                                })
                                        }else{
                                                $('#filtered-field').html('<div class="text-
center">No result</div>')
                                        }
                                }
                        },
                        complete:()=>{
                                $('.searched-user').click(function(){
                                        console.log('test')
                                        location.href =
"index.php?page=conversation&uid="+$(this).attr('data-id')
                                })
                        }
                })
        })
$('#convo-list li').click(function(){
        location.href = "index.php?page=conversation&convo_id="+$(this).attr('data-id')
})
websocket.onmessage = function(e){
                var data = JSON.parse(e.data)
                if(data == null)
                        return false;
                console.log(data)
                if(data.type=="chat_sent" && data.by != "<?php echo
$_SESSION['login_id'] ?>"){
                        if($('#convo-list li[data-id="'+data.convo_id+'"]').length > 0 ){
                                var li_clone = $('#convo-list li[data-
id="'+data.convo_id+'"]').clone()
                                 $('#convo-list li[data-id="'+data.convo_id+'"]').remove()
                                $('#convo-list').prepend(li_clone)
                        }else{
```

```
                        var li = $('#clone_ul li').clone()
                        li.find('img').attr('src',data.data.avatar)
                        li.find('.uname').html(data.data.name)
                        $('#convo-list').prepend(li)

                    }
                }
            }
            function renew_contacts(data){
                var users =  data.user_ids.split(',')
                var mid = '<?php echo $_SESSION['login_id'] ?>';

                if($.inArray(mid,users) > -1){
                    console.log($('#convo-list li[data-id="'+data.convo_id+'"]').length)
                    if($('#convo-list li[data-id="'+data.convo_id+'"]').length > 0 ){
                        var li_clone = $('#convo-list li[data-
id="'+data.convo_id+'"]').clone()
                         $('#convo-list li[data-id="'+data.convo_id+'"]').remove()
                        $('#convo-list').prepend(li_clone)
                    }else{
                        var li = $('#clone_ul li').clone()
                        li.attr('data-id',data.convo_id)
                        li.find('img').attr('src','assets/uploads/'+data.data.avatar)
                        li.find('.uname').html(data.data.name)
                        $('#convo-list').prepend(li)

                    }
                }
                $('#convo-list li').click(function(){
                    location.href =
"index.php?page=conversation&convo_id="+$(this).attr('data-id')
                })
            }
            function notify($convo_id=''){
                var li = $('#convo-list li[data-id="'+$convo_id+'"]')
                if(li.length > 0){
                    var notif = li.find('.notif').html();
                    notif = parseFloat(notif) + 1;
                    li.find('.notif').html(notif).show()

                }
            }
</script>
```

Appendix III

Conversation Monitoring Script

```
<?php
include 'db_connect.php' ;
	$convo_id =isset($_GET['convo_id']) ? $_GET['convo_id'] : '';
if(isset($_GET['uid'])){
	if(!isset($_GET['convo_id'])){
		$cdata = '' user_ids = '{$_GET['uid']},{$_SESSION['login_id']}' '';
		$cdata2 = '' user_ids = '{$_SESSION['login_id']},{$_GET['uid']}' '';

		}
	}
	$user = $conn->query("SELECT * FROM users where id =".$_GET['uid'])-
>fetch_array();
}else{
	$thread = $conn->query("SELECT * from thread where md5(id) =
'".$_GET['convo_id']."' ")->fetch_array();
	$user = $conn->query("SELECT * FROM users where id in
({$thread['user_ids']}) and id !=".$_SESSION['login_id'])->fetch_array();

}
?>
		<div class="w-100" id="cheader">
		<table class="" width="100%">
			<tr>
				<td width="10%"><div class="uavatar"><img
src="assets/uploads/<?php echo $user['avatar'] ?>" alt=""></div></td>

				</tr>
			</table>

		</div>

	<script type="text/javascript">
	var banned =
['MMM','XXX','xxx','mmm','https:','HTTPS:','Https:','http','Http','hTtp','HTtp','htTP','HTT
P:','give me your detail','Give Me Your Detail','GIVE ME YOUR DETAIL','Give me ur
detail','give me ur detail','Give me your detail','invest in this platform','Invest in this
platform','HTTP','INVEST IN THIS PLATFORM','Invest In This Platform','your account
detail','Your account detail','YOUR ACCOUNT DETAIL','GIVE ME UR
DETAIL','FUCK','Fuck','fuck','shit','SHIT','Shit','WWW.','www.'];
	document.getElementById('strip').addEventListener('keyup', function(e) {
		var txt = document.getElementById('strip').value;
		for (var x=0;x<banned.length;x++){
```

```
                        if(txt.search(banned[x]) !== -1){
                                document.getElementById('error').innerHTML = "<span
style='background: red; color: white' >"+banned[x]+"</span>" +" is detected as spam!";
                        }

                                        var regExp = new RegExp(banned[x]);
                                        txt = txt.replace(regExp, '');

                }

                document.getElementById('strip').value = txt;


                height: calc(10%);
                border-bottom: 3px solid #ececec ab;
                padding:1em .5em;
        }
        #cheader img{
                max-width: calc(100%);
                max-height:calc(100%);
                border-radius: 100%;
        }
                #cheader .uavatar {
        width: 50px;
        height: 50px;
        align-self: center;
        border-radius: 50%;
        border: 3px solid #808080c2;
        display: flex;
        justify-content: center;
        text-align: -webkit-auto;
        }
        #mbody {
        padding: 1em .5em;
        display: flex;
        align-items: center;
        justify-content: center !important;
                        overflow: auto;
        }
        #cbody{
                        height: calc(65%);
                                overflow: auto;
        }
        #cfooter{
                height: calc(40%);
                border-top: 3px solid #ececec ab;
```

```
        padding:1em .5em;
        position: relative;
        bottom: 0
}
.convo-right{
        padding:1em .5em;
        margin-left: calc(20%);
        direction: rtl;
        margin-bottom:5px;
        display: flex;
}
.convo-left{
        padding:1em .5em;
        margin-bottom:5px;
        display: flex;
}
.convo img{
        max-width: calc(100%);
border-radius: 100%;
        max-height:calc(100%);
}
        .convo .img {
width: 50px;
height: 50px;
align-self: center;
border-radius: 50%;
border: 3px solid #808080c2;
display: flex;
justify-content: center;
text-align: -webkit-auto;
}
.convo{
        position: relative;
        width: calc(80%);

}
.convo .message {
max-width: calc(80%);
border-radius: 15px;
        overflow-wrap: break-word;
}
.convo-left .message {
background: #a6a6af;
margin-left: 10px;
padding: 1em;
color: white;
```

```
        }
        .convo-right .message {
        background: #2031ff;
        margin-right: 10px;
        padding: 1em;
        color: white;
        }
        small.status {
position: absolute;
bottom: 0;
font-style: italic;
}
.convo-left .status {
left: 5rem;

        }
        .convo-right .status {
right: 5rem;

        }
        div#nw_msg {
                display: none;
position: absolute;
z-index: 10;
background: #00a1ff;
top: calc(79%);
right: 30px;
color: white;
padding: .5em;
border-radius: 10px;
cursor: pointer;
}
.message{
        direction:ltr !important
}
</style>
<script>
        $(document).ready(function(){
                load_convo()
        })

        function received_message(data,$convo_id = '<?php echo $convo_id ?>'){
                if($convo_id == data.convo_id){
                                var count = parseInt($('#cfield .convo').length) + 1;
                                chat_box(data.data,count);
                        }
```

```
	}
	$('[name="message"]').keypress(function(e){
		if(e.which == 13 && e.shiftKey == false){
			e.preventDefault()
			$('#send_chat').submit()
			return false
		}
	})
	$('#send_chat').submit(function(e){
		e.preventDefault()
		if($('[name="message"]').val() == '')
			return false;
		var data = {name:'<?php echo $_SESSION['login_name'] ?>',id:'<?php
echo $_SESSION['login_id'] ?>',avatar:'<?php echo
$_SESSION['login_avatar'] ?>',message:$('[name="message"]').val()}
		var count = parseInt($('#cfield .convo').length) + 1;
		chat_box(data,count)
		$('#cbody').animate({scrollTop: $('#mbody').height()},"fast");
		var frmData = $(this).serialize()
		$('[name="message"]').val('')



	websocket.send(JSON.stringify({type:'chat_sent',data:data,convo_id:resp.convo_i
d,by:'<?php echo $_SESSION['login_id'] ?>',user_ids:resp.convo_users}))
				}
			}
		})
	})
	window.chat_box=function(data,count,loaded =0){
		var ctype = '<?php echo $_SESSION['login_id'] ?>' == data.id ? 'convo-
right' : 'convo-left';
		var cfield = $('#convo_clone .convo').clone()
		cfield.addClass(ctype)
		var message = data.message
			message = message.replace(/\r?\n/g, '<br />')
		cfield.find('.message').html(message)
		cfield.find('img').attr('src','assets/uploads/'+data.avatar)
		cfield.find('img').attr('src','assets/uploads/'+data.avatar)
		cfield.find('.status').attr('data-id',count)
		if(loaded == 1)
		cfield.find('.status').html('Sent')
		if('<?php echo $_SESSION['login_id'] ?>' != data.id){
			cfield.find('.status').html(data.name)
		}
		var scrolltop =$('#cbody').scroll().get(0).scrollTop
```

```
                var scrollheight =$('#mbody').height()
                var scrolloff =parseFloat(scrolltop) +
parseFloat($('#cbody').scroll().get(0).offsetHeight)
                if(scrolloff >= scrollheight)
                        var autoscroll = true;
                else
                        var autoscroll = false;

                $('#cfield').append(cfield)
                if(loaded == 0){
                if(autoscroll == true)
                                $('#cbody').animate({scrollTop:
$('#mbody').height()},"fast");
                else
                        $('#nw_msg').show()
                }
                var li = $('#convo-list li[data-id="<?php echo $convo_id ?>"]')
                $.ajax({
                        url:'ajax.php?action=read_msg',
                        method:'POST',
                        data:{convo_id:"<?php echo $convo_id ?>",user_id:'<?php echo
isset($user) ? $user['id']: '' ?>'},
                        success:function(resp){
                                if(resp == 1){
                                        li.find('.notif').html('0').hide()
                                }
                        }
                })
        }
        $('#nw_msg').click(function(){
                                $('#cbody').animate({scrollTop:
$('#mbody').height()},"fast");
                                $(this).hide()
        })
        $('#cbody').scroll(function(e){
                if($(this).scrollTop() + $(this).innerHeight() >= $(this)[0].scrollHeight)
                        $('#nw_msg').hide()
                })
                                complete:function(){
                                $('#cbody').animate({scrollTop:
$('#mbody').height()},"fast");
                                }
                })

        }
</script>
```

Appendix IV

Login Script

```
<!DOCTYPE html>
<html lang="en">
<head>

                <div id="login-right">
                        <div class="card col-md-8">
                                <div class="card-body">

                                        <form id="login-form" >
                                                <div class="form-group">
                                                        <label for="username"
class="control-label">Username</label>
                                                        <input type="text" id="username"
name="username" class="form-control">
                                                </div>
                                                <div class="form-group">
                                                        <label for="password"
class="control-label">Password</label>
                                                        <input type="password"
id="password" name="password" class="form-control">
                                                </div>
                                                <div class="col-md-12">
                                                <button class="btn-sm btn-block btn-wave
col-md-5 btn-info float-right " type="button" id="create_account">Create
Account</button>
                                                <button class="btn-sm btn-block btn-wave
col-md-4 btn-primary float-right mr-2 mt-0">Login</button>

                                                </div>
                                        </form>
                                </div>
                        </div>
                </div>

</main>

<a href="#" class="back-to-top"><i class="icofont-simple-up"></i></a>

        $('#create_account').click(function(){
                uni_modal('Signup','signup.php')
        })
</script>
</html>
```

Appendix V

System Usability Scale Questionnaire

S/N	Questionnaire	Strongly Disagree	Disagree	Agree	Strongly Agree
1	This software seems to disrupt the way I normally like to chat.				
2	I would recommend this software to my colleague				
3	This software has at some time stopped unexpectedly.				
4	Learning to operate this software initially is full of problems.				
5	I sometimes don't know what to do next with this software.				
6	I enjoy the time I spend using this software.				
7	Working with this software is satisfying.				
8	The way that system information is presented is clear and understandable.				
9	The software documentation is very informative.				
10	Working with this software is mentally stimulating.				

Bio-data

Name: Oluranti John AFOLABI
Date of Birth: 30/08/1979
State of Origin: Ondo State
Nationality: Nigeria
Marital Status: Married
Address: SW9/388, Fatima, Odo-Ona, Ibadan.
Postal Address: P.M.B. 5382, NACGRAB, Moor Plantation, Ibadan.
E-mail Address: afolabi_oluranti@yahoo.com, olurantiafolabi50@gmail.com
Phone Number: +2348060738505

Career Objectives:
To apply practical knowledge and implement current technologies in attaining and maintaining high level of efficiency and effectiveness in the workplace and actualization of organizational objectives.

Educational Institutions Attended with Dates:

a) Ladoke Akintola University of Technology, Ogbomoso 2013 – 2014
b) The Polytechnic, Ibadan 2006 – 2008
c) Osun State College of Technology, Esa-Oke 2002 - 2004
d) Apata Community Grammar School, Ibadan 1992 - 1997

Certificate Obtained with Dates:

a) Post Graduate Diploma (PGD) Certificate (Computer Science) 2014
b) National Youth Service Corps (NYSC) Certificate 2010
c) Higher National Diploma Certificate (Computer Science) 2008
d) National Diploma Certificate (Computer Science) 2004
e) National Examination Council Certificate 2003
f) Secondary School Certificate 1997

Work Experience with Date:
Assistant Chief Data Processing Officer (NACGRAB) 2005 till date

Responsibilities:

a) Web Developer/Administrator
b) UX/UI Designer
c) Database Manager
d) Variety Release Database/Catalogue Manager

Training/Workshop Attended with Date:

1. Training on Internet/Web Management – May 2008 at National Centre for Technology Management (NACETEM), Lagos, Nigeria
2. 6 Months Training on Web Development (Dreamweaver, HTML, CSS3, MySQL, PHP, eCommerce, Context Management System (CMS) platforms) - Don Dada Technology, Ibadan, Nigeria - 2010
3. DUS and VCU Training by representatives of IOWA State University, USA – January 27th – 28th 2011 at NACGRAB, Nigeria

4. Training on Seed Policy Enhancement in African Region (SPEAR) by representatives of IOWA State University, USA) – June 18th 2011 at NACGRAB, Nigeria
5. Training workshop on implementation of Seed Policy Enhancement in African Region (SPEAR) – September 26th – 28th 2011 at National Agric Seed Council (NASC), Abuja, Nigeria
6. Regional Training Workshop on implementation of the ECOWAS Seed Regulation – February 24th to 1st March, 2014 – Abuja, Nigeria
7. National Seed Planning meeting and ECOWAS Seed Regulation Awareness Workshop – June 9th – 11th 2014 – Abuja, Nigeria
8. Workshop/Training on the Regional Variety Release Management System/Catalogue – May 11th – 22nd, 2015 – Accra, Ghana.
9. Workshop/Training on the Regional Variety Release Management System/Catalogue – February 15th – 19th, 2016 – Saly, Senegal.
10. Workshop/Training on the Regional Variety Release Management System/Catalogue – February 22nd – 26th, 2016 – Banjul, The Gambia.
11. Workshop/Training on the Regional Variety Release Management System/Catalogue – March 21st – 25th, 2016 – Abijan, Cote d'lvoire.
12. Regional Workshop on Regional Catalogue of Plant Species and Varieties – April 4th–8th, 2016 – Bamako, Mali.
13. Regional Workshop on Regional Catalogue of Plant Species and Varieties – May 23rd–27th, 2016 – Dakar, Senegal.
14. Training on and Launching of the Electronic Seed Platform (WASIX) – July 28th – 29th, 2016 – Denis Hotel, Abuja.
15. Finalization of the Regional Catalogue of Plant Species and Varieties – January 16th – 20th, 2017 – Dakar, Senegal.
16. Government Wide Messaging and Collaboration (GWMC) Email Configuration – April 19th – 20th, 2017 – Galaxy Backbone Office, Wuse, Abuja
17. Updating Regional Catalogue of Plant Species and Varieties – Feb 23rd – 27th, 2018 – Dakar, Senegal.
18. Training on Seed Germplasm Management System (Genesys), IITA, Ibadan, Nigeria, October, 2019
19. 2nd International Conference on Applied ICT (ICAICT), Lead City University, Ibadan, Nigeria. Theme: ICT for All. ICAICT 2019.
20. Updating Regional Catalogue of Plant Species and Varieties – Aug. 26th – 30th, 2019 – Ouagadougou, Burkina Faso.
21. Online HTML Fundamental course with Sololearn – August 26th, 2020
22. Online JavaScript Tutorial course with Sololearn – September 2nd, 2020
23. Online PHP Tutorial course with Sololearn – September 7th, 2020
24. Online SQL Fundamental course with Sololearn – September 8th, 2020
25. Technical meeting for the development of the terms of reference and the roadmap for updating the electronic data platform of the regional seed and seedling catalogue of West Africa. – February 2nd – 5th, 2021 – Ouagadougou, Burkina Faso.

26. Updating Regional Catalogue of Plant Species and Varieties – June 7th – 11th, 2021 – Bamako, Mali.
27. Updating the Electronic platform of plant species and varieties in West Africa and the Sahel – February 28th – March 4th, 2022 – Bamako, Mali.
28. Genebank Data Workshop by Global Crop Diversity Trust and IITA – April 25th – 28th, 2022 – IITA, Ibadan, Nigeria.
29. Seeds for Resilience GOAL Workshop by Global Crop Diversity Trust – May 9th – 13th, 2022 – Nairobi, Kenya.
30. Training on the Electronic Platform of Plant Species and Varieties in West Africa and the Sahel – May 30th – June 2nd, 2022 – Bamako, Mali.

Co-Curricular Activities:
Reading, Learning, Driving, Meeting people, Travelling & Singing

Additional Information:

- A team player with pleasant disposition who can work with little or no supervision
- Ability to work under pressure
- A calm approach when working in tense situations
- Drive for excellent and strong appetite for knowledge
- Great time management skills
- Effective communication and presentation skills.

Names and Address of Referees:

Prof. Phillip Achimugu
Associate Professor
Computer Science
Air Force Institute of Technology, Kaduna, Nigeria.
check4philo@gmail.com
+2348072309773

Dr. Sunday Aladele
Director/Chief Executive Officer
National Centre for Genetic Resources and Biotechnology (NACGRAB)
Moor Plantation, Ibadan, Oyo State, Nigeria
saladele6083@gmail.com
+2348038074937

Dr. Folarin Okelola
Technical Adviser to the Director General
National Agric Seed Council (NASC), Abuja, Nigeria.
fspkelola@yahoo.com, fspkelola@gmail.com
+2347036046157

Signature	**Date**

University Compliance Certificate

This is to certify that this thesis by Afolabi, Oluranti John with Matric No LCU/PG/000141 in the Department of Computer Science, Faculty of Natural and Applied Science, Lead City University, Ibadan, is in FULL compliance with the approved university format and style.

Signature	**Date**