Project Dissertation Report on

Behavioral Response of Individuals to Cyber Attacks

Submitted By

Karan Chaudhary
Roll No: 2K19/DMBA/043

Under the Guidance of

Professor Yashdeep Singh
Delhi School of Management

DELHI SCHOOL OF MANAGEMENT

Delhi Technological University
Bawana Road Delhi 110042

Jan-May 2021

CERTIFICATE

This is to certify that the work titled **'Behavioral Response of Individuals to Cyber Attacks'** as part of the final year Major Research Project submitted by Karan Chaudhary in the 4th Semester of MBA, Delhi School of Management, Delhi Technological University during January-May 2021 was conducted under my guidance and supervision.

This work is his original work to the best of my knowledge and has not been submitted anywhere else for the award of any credits/ degree whatsoever.

The project is submitted to Delhi School of Management, Delhi Technological University in partial fulfillment of the requirement for the award of the degree of Master of Business Administration.

Signature of Project Guide
Prof. Yashdeep Singh
Delhi School of Management
Delhi Technological University

Signature of Head
Prof. Archana Singh
Delhi School of Management
Delhi Technological University

DECLARATION

I hereby declare that the work titled **'Behavioral Response of Individuals to Cyber Attacks'** as part of the final year Major Research Project submitted by me in the 4th Semester in MBA, Delhi School of Management, Delhi Technological University, during January-May 2021 under the guidance of Prof. Yashdeep Singh, is my original work and has not been submitted anywhere else.

The report has been written by me in my own words and not copied from elsewhere. Anything that appears in this report which is not my original work has been duly and appropriately referred/ cited/ acknowledged.

Karan Chaudhary
2K19/DMBA/043
MBA (Operations and IT)

ACKNOWLEDGEMENT

It is a great pleasure for me to acknowledge the kind of help and guidance received during the research work. I would like to thank my faculty advisor Prof. Yashdeep Singh, who helped me to take up the topic **'Behavioral Response of Individuals to Cyber Attacks'** and guided me to complete this project properly. The project provided me with an excellent opportunity to explore the areas of Data Science, Cyber Security and Analytics.

I am highly indebted to Delhi School of Management, Delhi Technological University for giving me an opportunity to work on this project. Lastly, I would like to express my gratitude to all the honorable faculty members for sharing their experience and expertise on this Project.

I have put all my efforts to ensure that the project is completed in the best possible manner and ensured that the project is error-free.

Karan Chaudhary

(Roll No. 2K19/DMBA/043)

EXECUTIVE SUMMARY

The idea of the present research "**Behavioral Response of Individuals to Cyber Attacks**" came when we began to come across the stories of cyber-attacks as a part of daily routine. We as an individual might not even know how vulnerable our important data is to the outside world and what data has already been leaked off our system while we just relax and without worry carry our work, tasks, scrooge over some entertainment stuff, e.g., games, movies, songs, etc. and within a nick of a second, we might get ourselves entrapped into the clever pitfall sewn by hacking agencies or agencies with malicious intent. The first time this idea of taking "**Behavioral Response of Individuals to Cyber Attacks**" when we heard about the troublesome Ransomware attacks that basically took over the control of millions of user's systems or workstations imposing a virtual lock on the top of file or directory system disabling access unless user pays heavy ransom to unlock the lock and get their data back.

This attack is just a sand grain in the dessert and there are myriad other examples that revolves around the news almost every single day wherein attacks of this nature can jeopardize the victims' data and their identity and even critical data stored in the system, such as, the user session logged in to any personal account may get freeze into the hands of cyber attacker at the expanse of attacker's advantage (and the attacker may even change the credentials of user's account).

Through this research paper, we do not intend to algorithmically and with crypto science solve the cyberattacks against their digital signature and come up with techniques that antivirus engine deploys in its software to potentially block any breach, but we will target the user's perception based on impact the attack has left in an individual's day-to-day life of their engagement with the system. The paper is meant to capture the sentiment of the user as in the effect of the cause in user's mind and the immediate response of user towards the entity wherein they think the root of the problem lies, the user's perception about the system or platform when the attack happens, the probable cause (or entity) upon which user's displays their anger as an immediate response as per their own assumption where things might have gone wrong and the emotional behavior exhibited unbeknownst to user's knowledge when the attack leave their data or system exposed.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1
Introduction

This project titled “BEHAVIORAL RESPONSE OF INDIVIDUALS TO CYBER ATTACKS” is a **qualitative research** report based on secondary data collected, recorded, cleaned, measured, analyzed, and visualized to understand the user’s mindset. Secondary data in this case is tweets collected from Twitter, a social media platform.

Attackers are becoming more common in the history of cyber-attacks and are showing new intent through advanced cyber-attacks. Unfortunately, cybercriminals use anonymity to search for profitable business models on the Internet. Serious situation that needs to be corrected by network defenders. Thus, the effectiveness of modern methods and practices requires a paradigm shift. Given that most cyber security incidents are man-made, these changes will require the extension of research to undiscovered areas such as aspects of cybersecurity behavior. To improve the current situation, it is important to focus on social and behavioral issues. This White Paper outlines relevant theories and principles and provides information that provides a comprehensive framework that combines cyber-behavioral security, human factors, modeling, and simulation.

With the increasing access to internet, online platforms like Facebook and Twitter have become increasingly common where people express their views on different topics. Twitter, being a common platform, it gives people freedom of speech to express their opinion. Therefore, we will use twitter sentiment analysis in our case study to look for the recent cyber-attacks that took place in a user’s system through tweets expressing their concern and emotions through tweets.

Twitter, a microblogging site, allows you to express your opinion in a few lines rather than the story, making it relatively easy to find textual content compared to other social networking platforms. Twitter is a U.S. microblogging and social networking service where customers post and manage the "Tweet" message. Initially, tweets were limited to 140 characters, but in November 2017, the character limit was reduced to 280 characters, which doubled in non-Asian languages. As of May 2019, 152 million active users have tweeted with 500 million a day.

Sample Tweets-

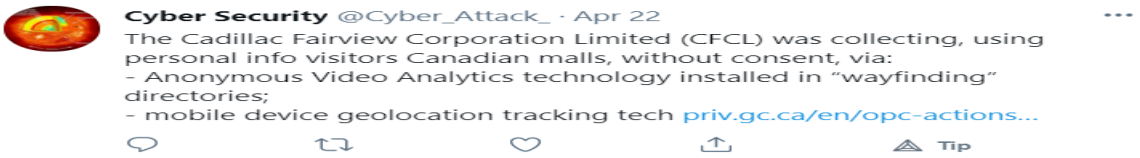

Figure 1: Sample Tweets from Twitter Website
***Source**: Taken from own Twitter account*

The below graph gives graph-wise distribution of few of the most common and major threats and malwares that are experienced by organizations, servers, systems, and user-systems across the world. From the graph, Denial-of-Service attacks accounts for the greatest number of cases.

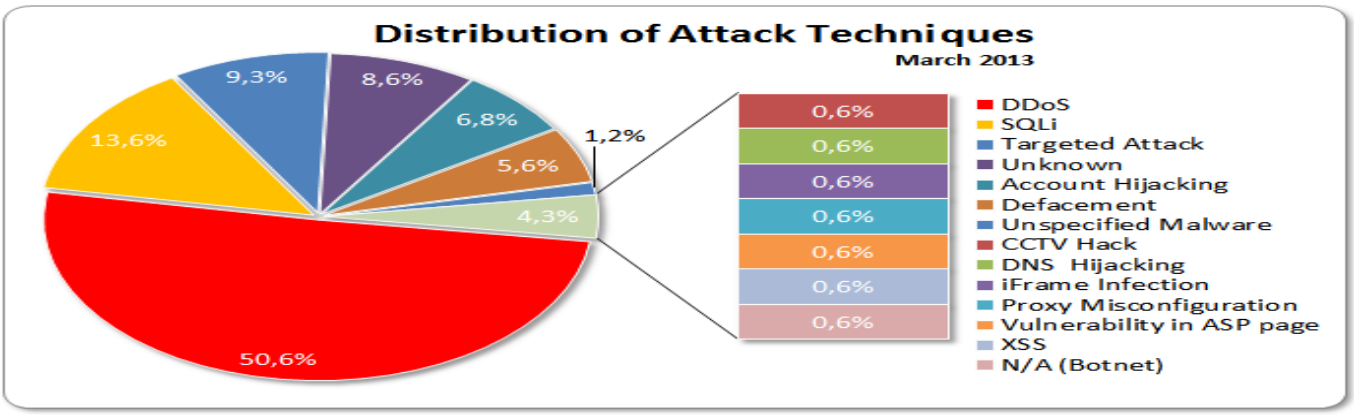

Figure 2: Few topmost attacks that happen in a common system
Source: *https://www.cyberintelligence.in/top-ten-countries-with-weak-cyber-security/*

Twitter Sentiment Analysis is he process of defining the emotional tone behind words that are used to understand the attitudes, thoughts, and feelings expressed in comments made on the Internet.

This is especially useful when monitoring Social media. This gives a picture of the public on a particular topic. As well as; it is also convenient to use in situations where business intelligence and text analysis are required.

The research paper uses an Application Programming Interface (API) service to collect data from the Twitter Social Network. The second step is to process the collected data. This is done using a software library to edit the natural language. The data collected are analyzed using methods and classifications to get an idea of the sources and individuals who create the data. The next step is to analyze the availability of unstructured data, which is mainly text description using text analysis.

The first chapter of the Introduction contains several sections, including the need of the study, Statement of the problem, Objectives of the study, Scope of the Study, and organization of the Study.

1.0 Background

The present Research on "**Behavioral Response of Individuals to Cyber Attacks**" is done as a part Major Research Project in the final semester (IV) of Management course (MBA). All the work from beginning to end in this report had been done during the online mode of classes through the help of online research and research papers on Cyber Security attacks trending on twitter. Some of the help is taken from friends and relatives who have been knowingly or unknowingly have fallen to the trap of these attacks.

The survey analysis of Global Cybersecurity threats indicate that data breach and cyber theft incidents have increased multifold during the COVID-19 period. According to the Internet Security Threat Report (Symantec 2017), the average ransom was $ 373 in 2014 and $ 294 in 2015. It returned to $ 10.77 in 2016, but it can be assumed that this is due to the growing value of bitcoin. The benefits of making money and low-risk offenses encourage cybercriminals.

Attackers (Hackers) are becoming more common in the history of cyber-attacks and are showing new intent through advanced cyber-attacks. Unfortunately, cybercriminals use anonymity to search for profitable business models on the Internet. Serious situation that needs to be corrected by network defenders. Thus, the effectiveness of modern methods and practices requires a paradigm shift. Given that most cyber security incidents are man-made, these changes will require the extension of research to undiscovered areas such as aspects of cybersecurity behavior. To improve the current situation, it is important to focus on social and behavioral issues. This White Paper outlines relevant theories and principles and provides information that provides a comprehensive framework that combines cyber-behavioral security, human factors, modeling, and simulation.

Digital currency is preferred to extortion because it is universally accepted without revealing a user's identity. The same report shows that the number of designated extortion programs in 2016 increased by 463,841. In addition, more than 7.1 billion personal data have been compromised over the past eight years because of cyber-attacks. The number of malware attacks is increasing. For example, repeating malicious code that removes the "Shamoon" disk in the Middle East, or computer attacks targeting Ukraine involving the Kill Disk Trojan horse. Citing the example of a Ukrainian power system attacked by a cyber-attack in May, it shut down about 225,000 customers, which was used to remove the main boot file and logs of the organization's modified Kill Disk target system. It is used to enhance attacks caused by stations, servers, and human-machine interface maps. Trojan horses are considered a third wave of malicious code that spreads through malicious websites or emails on the Internet. Of course, data breaches are one of the most devastating cyber-attacks. Figure 3 depicts three main cyber targets, or their combination.

They are usually referred to as CIA triad:

- Confidentiality threats (data theft) that can target database administrators, backup servers, and administrators.

- Threats to integrity (data alteration) include theft, misleading financial data, stealing large sums, redirecting direct deposits, and damaging an organization's image.

- Accessibility attacks (denial of access) can be distributed in the event of a

distributed denial of service (DDoS) targeted denial of service and physical disaster.

After reaching the first level of the network, an attacker will try to enter any level of the security system. So, attackers need to be motivated to analyze security at all levels with security vulnerability detection tools before doing so. Black Report 2018 pays special attention to both the piracy phase and the period when intruders, according to the industry, both invade the organization's cyber systems. Most respondents said they could access an organization's systems within 15 hours, compare valuable data to find and map it. Currently, according to most industry reports, the average time to detect an intrusion is 200-300 days.

Clearly, cybercriminals or crimes still have an advantage over cyber-defenders. What are the weaknesses of current research and in which areas do you need immediate attention and improvement? Further research on cybersecurity behavior is needed, and we believe that integration with human factors and the application of advanced modeling and simulation techniques can accelerate developments. Our research highlights two important aspects.

(1) A holistic approach to cybersecurity is important and should be defined in terms of understanding cyberspace. We agree with the definition of an international organization for the standardization of cyberspace. "It's a complex environment that results from the interaction of people, software, and services using technology devices and networks connected to the Internet and doesn't exist in physical form." This definition presents cyberspace as a complex environment and interacts with people. Thus, people's prejudices and behaviors influence their interactions with software and technologies that affect the virtual space. We believe that conducting this interdisciplinary study could increase the relevance and number of handwritten cybercrimes in key journals. Some cybercrime manuscripts have observed those dealing with cybercrime. So, let us look at some theories of behavior and crime. According to the proposed interdisciplinary approach, the cyber group should include people from different backgrounds, such as IT, criminological, psychological, and human factors.

(2) Businesses need to consider vulnerabilities, including human factors, when designing systems. Fixing a vulnerability is a much better solution than spending resources to fix or protect it. This may seem like a trivial proposition, but many supporters and users often see security as a minor issue if security is

not the primary function. Security - This is the primary responsibility of users of the information infrastructure. In addition, system developers focus on user needs before integrating them into the security architecture. Over time, security devices will emerge that are easy to manage or meet different system requirements. This explains our view that modeling and simulation are important factors. Stakeholders such as users, administrators and developers need to be involved in building this model, in particular the model for assessing cognitive load and threat response time. Stakeholders can test real-life scenarios of social engineering attacks with simulations. In addition, the description of the vulnerability may be affected by the budget. Businesses are minimizing their cybersecurity budgets. Financial institutions spend an average of 10% of their cyber security IT costs and an average of 0.3% of revenue. Nowadays, some businesses spend more on cybersecurity, but this is an area that may not provide maximum security. Organizations spend more on security, but that is not smart. These are called post-insurance costs and lead to widespread inefficiencies. Of course, this situation adds to the complexity of the security issue. Thus, different industries have different thinking about their cyber security needs and in most cases do not even exist.

Part of the search command is: (Cyber Security and Human Agents), (Cyber Security and Behavioral Aspects), (Cyber Security and Modeling and Simulation) (Interdisciplinary Approach and Cyber Security) (Cyber Security), and Criminal Theory). The end user is recognized as having a key for the backdoor of the network and is a critical element in determining the entering into the system. The behavioral science method is used to identify the factors that influence user behavior in the field of cybersecurity. Security awareness and common external factors influence a person’s adaptive behavior in cybersecurity. These factors depend on the characteristics of the user (gender, age) and the work environment. Some theories of criminology provide an important basis for the empirical study of cyberspace nodes in the criminal ecosystem. Criminologists are still unable to put forensic science at the forefront of criminology. The most common criminological explanations for cybercrime are learning theory, self-control theory, neutralization theory, and everyday theory. Perhaps because some criminologists deal with cybercrime, joining a cybercrime system is not quick. The two themes on the action page fall into the following categories: (1) Cognitive burden that can cause careless blindness. This way, team members cannot feel the unexpected events while concentrating on the main task. (2) Prejudice. It supports architects and security developers. Wait for the findings and examine them in the project.

Figure 3: Losses caused by Cyber Threats

***Source**: https://www.nature.com/articles/d41586-020-00758-2*

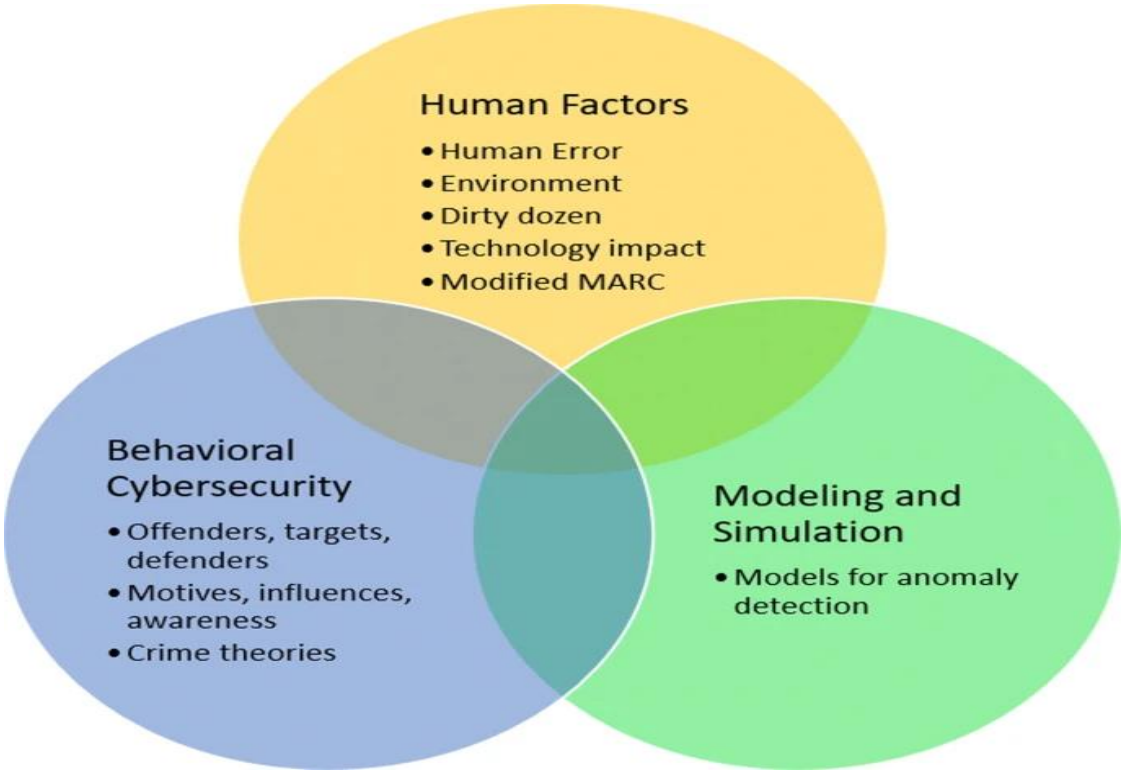

Figure 4: Venn Diagram for the interdisciplinary framework

Source: *https://www.kaspersky.com/resource-center/definitions/what-is-cyber-security*

Following figure summarizes and group all user errors and the insider into human error from the Venn diagram represented before: -

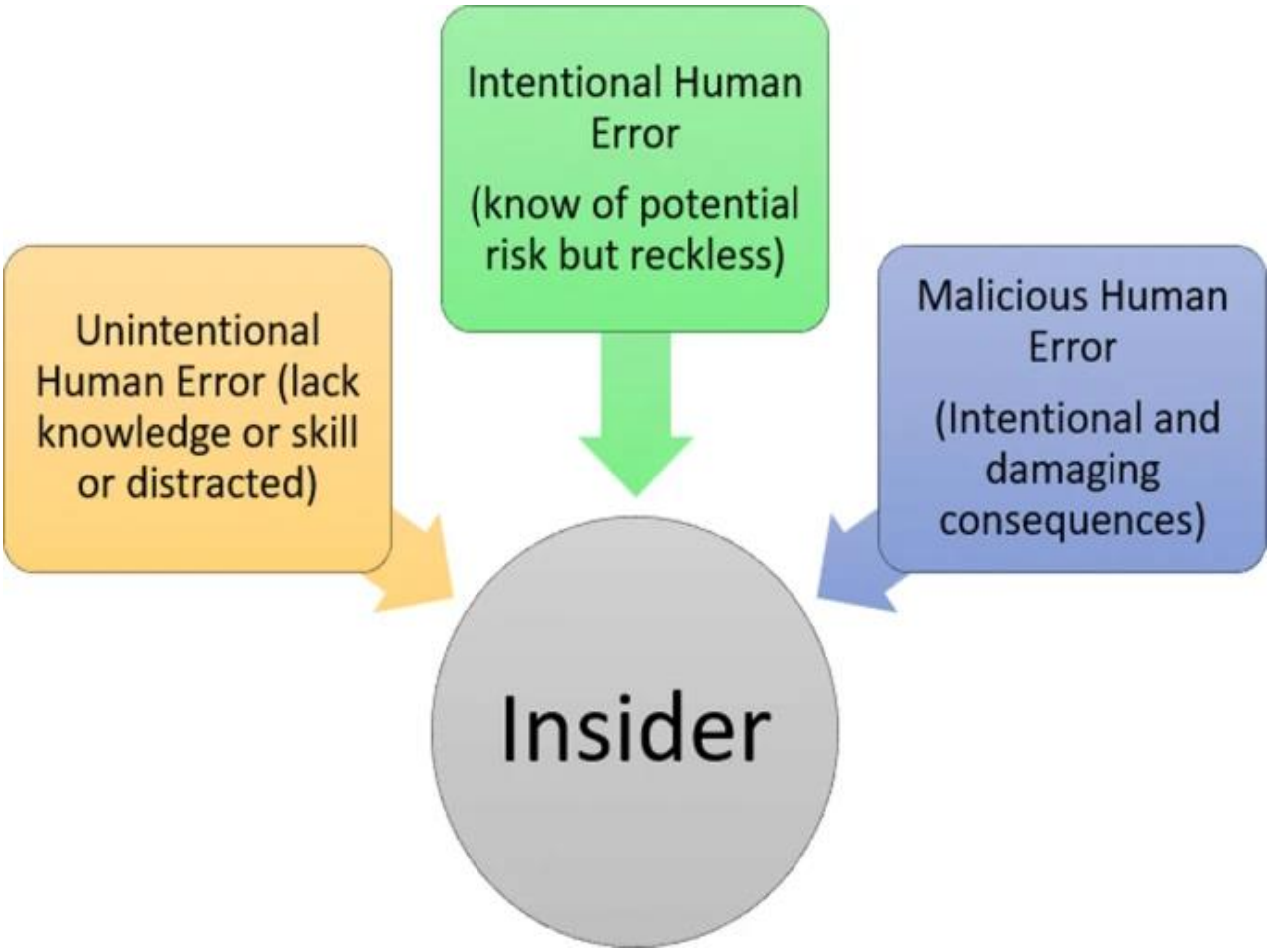

Figure 5: Proposed UIM human error as insider-anomaly concept

Source: *https://digitalguardian.com/blog/what-cyber-security*

Therefore, research is consistent in recognizing that behavioral patterns have not yet been adequately studied and the focus is on the technical side. One problem is the complexity of the model when we examine different theories. Our goal is to give you an idea of the current problem. For example, if you classify an internal threat as a human error, the internal problem is a design requirement. This understanding makes our approach essential by opening channels for the application of best practices by human factors such as healthcare, aviation, and the chemical industry. This reinforces the idea of an interior that has a design requirement (prevention).

1.1 Motivation

Not all partial solutions (firewall, IDS / IPS, network process, proxy, email gateway, etc.) will be the perfect solution, so offenders will still have maximum choice at the network level, so they will have to invest. The Full frame interdisciplinary structure serves as a background for understanding the relationships between relationships and for improving the study and maturity of the security program. The Venn diagram shows three areas:

- The focus of our research is on the behavioral security of cyberspace. We examine the profile and behavioral methods of internal hackers, the theory of social crime. They also recognize influential weapons that are often overlooked by the defense used by criminals.
- Integrating Human Discipline into Cyber Security Behavior We are talking about human factors that lead to human error. By treating internal problems as human error and improving the environment, you can reduce risks and adapt them to future system design requirements. The internal hazards of existing vulnerabilities or conditions should be considered when designing the system. The U.S. National Institute of Standards and Technology (NIST) recommends that the best way to engage everyone is to use cybernetic incentives to move everyone around. Therefore, human factors need to be integrated to improve the work environment, reduce risk, and reduce the likelihood of system failure.
- We recommend the use of models and simulations to explore, develop, and implement new methods, tools, and strategies. Simulation and simulation are useful for a variety of reasons and can be extended to situations such as when the actual experiment is annoying, dangerous, or low. The simulation can test the use of human agents, such as whether the actual process can prevent the end user of the security system from losing sensitive information and cognitive threats. We study modeling and simulation in the literature and provide information based on attention to human error.

Behavioral security in cyberspace is important, and there is no doubt that further research is needed. As human productivity is not only influenced by education, which is the main activity of cybersecurity, it highlights three elements of the proposed interdisciplinary structure. This is since the system itself is also influenced by human prejudices, load management, communication methods, human computer interfaces, traditional distractions, and so on. Several factors contribute to delays in research and the interdisciplinary approach.

Unfortunately, many companies underestimate the severity of computer incidents or shift the responsibility to the individual when a problem arises. For example, the Federal Trade Commission website reported a data breach in September 2017, where Equifax provided 147 million personal information and Equifax entered into a global agreement with the Federal Trade Commission, the Financial Protection Agency and 50 states and the territories of the United States. The deal includes up to $ 425 million to help those affected by privacy. However, the conciliation has little effect on the person making the claim (a lump sum of $ 125 or credit supervision over several years). Individuals cannot deny that Equifax will be the data controller. This has strained many people. According to most online reports, Equifax does not update known vulnerabilities in Apache Struts Web application software. However, on October 3, 2017, the CEO of Equifax informed lawmakers that the mass violation was due to an employee's fault.

1.2 Need of the Study

Cybercrime Offenders: Hackers

Hackers' techniques

A hacker is a person who gains unauthorized access to data to modify, delete or sell it using technical intelligence. Hackers can take a variety of steps to succeed, but common network interventions include gathering information, accessing vulnerability profiles through a scan, or hacking an access point or level, accessing support for another level, or installing a program. You can hide tracks by making them accessible and deleting them. Invasion techniques can be classified as follows.

- Dictionary attack to crack weak passwords. It is like a brutal attack for security. This takes advantage of the fact that the user remembers complicated or meaningless passwords, so use a relevant simple password. Hackers can find users with weak passwords (such as 123456 and passwords). Companies correct passwords and order specific change processes. However, users still use the same password on different sites.
- It changes the structure of SQL queries by inserting malicious code into the structured query (SQL) language. Manages the location database.
- Cross-Site Scripting (XSS) is an attacker who implements a malicious script on a victim's website.
- Phishing is a social engineering attack in which phishing deceives users into disclosing confidential information.
- Failure of some networks Violation of the wireless network. The access point and default password of these network providers are not changed. If your Wi-Fi network has a vulnerable access point, there is a possibility of violation. Hackers use to scan and list ports.

- Keylogger is a program that runs in the background and records user keystrokes. This allows hackers to record their credentials.

The profiles of some hackers are described in literature reviews. They have different educational qualifications, many qualifications, no copyright work, or work for organizations. Hackers can be beginners in the scenario. Their goal is curiosity and popularity. Cyberpunk, like virus makers, have average capabilities and their intentions can share some financial gains. An internal body that is confidential or above can be governed by various motives, such as revenge and monetary interests. Internal skills are generally high. It is strange or popular that a gray hacker intends to write a little thief virus or a former guard hacker with a high level of skill. The motives of professional criminals or black-hat hackers can be monetary and have an extremely high chance. The motives of the information warriors, the mercenaries of cyberspace, are mainly spies, they belong to the nation-state group. Political activists and Hacktivist are ideologically motivated and may include highly educated members in some respects.

Insights on Hackers' Techniques

It is important to understand these piracy techniques and the hacker's motives and to predict the hacker's behavior. Not all hackers think of themselves as defenders or straight men. Defenders therefore need to be interdisciplinary to consider different skills and fight. This hypothesis is confirmed by one of the true stories of the hacker. - Hackers changed the slot machine's firmware after hiring an employee in the casino or casino. Their motive was money, and the motive was that mechanical programmers were human, so they probably had a back door to the program. Hackers checked the Patent Office because they had to enter the code to apply for a patent. The secret was revealed by code analysis. The machine is a pseudo-random random number generator A 32-bit random number generator that could be easily interrupted. System developers do not want to create real random numbers, so they have some control over the options and the game. The hackers in this story are programmers, and their ideas were simple enough to find instructions to achieve their goals. We are currently spending money on security, not on finding casino security resources. The hacker said he did not even regret being robbed by the casino. Casinos were stolen by people.

So here are some questions that need to be answered regularly to predict the next steps for hackers. Have the attack surfaces been defined? The target area contains the sum of all attack vectors that a hacker could attempt to exploit the vulnerability. What is the most vulnerable or destructive tool when exploited? How are access points protected? How can intruders get into the crown jewel? Examples of crown jewelry are the most valuable data. Where is the hoop jewelry (servers, networks, backups, etc.)? Do you know the list of allowed and unlicensed devices? Is the operating system properly configured and updated? Do you have control over stolen credentials or hacked user accounts? What malware protection are used? How effective are training and awareness programs? Are your employees aware of the dangers of social media? How do employees work in a production environment? How effective and reliable is the intrusion detection system? Is the reporting system for potential

threats or violations clear? Are there any plans to combat the internal threat? It should be noted that many companies recognize that focusing on prevention increases costs and reduces productivity. This increase is due to responses to incidents and interactions with safety management systems. Poor performance is caused by authentication or re-authentication of credentials or user accounts. They need to analyze the costs of different options, including prevention programs, event management programs, or hybrid options.

Cybercrime offenders: Insiders

Insiders' threat

Insiders are organizational hackers. So confidential have power behind a firewall. Internal threats have been widely recognized as a key issue in cybersecurity management. Businesses were concerned about accidental data breaches caused by threats, not malicious data breaches. However, their fears do not necessarily lead to effective change in computer programs. According to SANS Healthcare's cybersecurity survey, 51% see negligible trust as a significant threat to human cybersecurity behavior. Many theories can also be applied to patterns of behavior that serve to understand risks and motivations. Management policies and guidelines often focus on cybercriminals, but the rationale for users and supporters is the vulnerability of cyber systems. Irrational behavior is likely to be dangerous and unpredictable, based on frustration, anger, and dissatisfaction with work. Computer protectors often do not control irrational behavior. End-user behavior in an organization is likely to be attributable to these groups of behaviors and lead to intentional harm, harmful abuse, dangerous interventions, unintentional errors, conscious trust, simple hygiene, and the use of intentional and technical knowledge as a model. Bots can be used to abuse privileges. An insider threat management (MERIT) management model that can be used to report threats to individuals. After analyzing some cases of internal computer disorder, 7 observations were checked and confirmed. These observations showed that people who survived stressful events (sanctions) that were generally confidential and dissatisfied with employees, behavioral tendencies (drug use, aggression, etc.) as well as physical and electronic access that had unknown channels created for post-dismissal attacks. (Insufficient access works). The limitation of the investigation into the threat of confidential information is the lack of data.

1.3 Statement of Problem

In many cases, when a user sends a request to a competent authority, that is, a security service subscribed by the user, the members of the organization do not understand the seriousness of the situation and cannot return the user lost data (or query) in the attack. Because of their failure, they cannot put trust in the customers or security services subscribers thus creating a social and economic impact on the organization. This study sought to identify these problems.

1.4 Objective of the Study

The purpose of the study is to process the tweets and people behavioral response and identify the sentiments of the people towards their security services, Systems, and overall emotional blow during the theft of identity or privileges or data (Cyber theft for this project) and to observe whether those sentiments are helpful in dealing with issues by the organizations or regain trust and confidence on services. Through this analysis we can determine the type of response a company should be prepared with in advance in the event any unforeseen thing happens and the action on behalf of the user's lose by the company to not just deal with the source of the attack, but to provide irredeemable justice to the victim. So, for this I have chosen topmost cyber issues and attacks and tried to identify the sentiment of people towards them at the time of any theft.

Objectives are: -

1. To do the Sentiment Analysis and observe the sentiments of people towards their services and system during and after the time of cyber-attack.
2. Identify the possible reasons for those sentiments by doing Text Mining.
3. To observe the behavioral response in response to the theft and observe the cause that leads to that effect.
4. To devise possible crisis response plan for organizations user relies upon.

1.5 Scope of the Study

The scope of the research is limited to sentiment analysis using Twitter (Tweet). R Studio was used to extract texts and sentiment. The results are analyzed, and accordingly resolution mechanism is suggested.

The usefulness of the project is that it provides a complete reference for analyzing Twitter data. It can be generalized as: -

1. Use R to export data, create a developer account, and access tweets using keys and distinctive APIs.
2. Sentiment analysis by receiving many tweets using the syuzhet library.
3. Intelligent text analysis to find the most used words in the source data.

1.6 Organization of the Study

The present report has been organized as follows. There are further 5 more chapters. They are Literature Review, Research Methodology, and approaches in which we will discuss detail types and techniques of sentiment analysis, Data Analysis, Limitations and Conclusion and Recommendations.

Chapter 2
Review of Literature

2.0 Introduction

This section describes the relevant work in the field of sentiment analysis. Many companies already use this sentiment analysis method to identify customer opinions about their products, brands, or services. It is a widely used and accepted method of marketing, financing, insurance, etc. Sentiment analysis helps companies create new products or improve existing products by judging people's demand or satisfaction.

Sentiment analysis is especially useful in political scenarios, such as when the government wants to know how satisfied our compatriots are with their current policies and actions. You can decide if people trust their leader and agree with the decision. This has proven to be extremely helpful if you want to get to know people's attitudes towards the leader and take the necessary action accordingly.

This section provides a general overview of the Sentiment Analysis, Related work done on this field and limitations of prior Art.

2.1 General Sentiment Analysis

Sentiment analysis algorithms and applications: A survey (Medhat, Hassan, & Korashy, 2014)

This review provides a comprehensive assessment of past developments in this area. In this review, several recent developments in the proposed algorithm and various SA (sentiment analysis) are easily explored and explained. This document is categorized according to their contribution to the various SA strategies. It describes the areas related to transition learning, emotion discovery, and resource creation (SA) that attract the latest researchers. The main purpose of this review is to quickly explore the almost complete holistic details of the area covered by the SA method. The main contribution of the article is a complex classification of recent articles and current trends in mood analysis and research in related areas.

Sentiment Analysis and Opinion Mining (Liu, 2012)

Because this research is important to business and society, it ranges from computer science to business and social science. The importance of sentiment surveys is consistent with the explosion in social networks such as reviews, forums, blogs, microblogs, Twitter, and social networks. This book is a comprehensive review of the original text and content. More than 400 links cover all major topics and current events in the topic. Suitable for students, researchers, and professionals interested in evaluating the benefits of social media, particularly by analyzing attitudes. In exceptional circumstances, educators can conveniently use them in natural language processing, social media analysis, textual content analysis, and statistical analysis courses.

Lexicon-Based Methods for Sentiment Analysis (Taboada, Brooke, Tofiloski, Voll, & Stede, 2011)

The author describes a vocabulary-based method for extracting keys from text. Semantic Orientation Calculator (SO-CAL) Semantic Orientation (Polarity and Volume) Includes amplification and denials using a dictionary of annotated terms. SO-CAL is used to determine the type of polarity, that is, how to assign a positive or negative label to text. This reflects the opinion of the text on the key issues. The authors show that the overall performance of SO-CAL is stable in terms of domain names and completely hidden data. It also explains how to create dictionaries and how to use Mechanical Turk to check the consistency and reliability of dictionaries.

Opinion Mining and Sentiment Analysis (Pang & Lee, Opinion Mining and Sentiment Analysis, 2008)

The increasing availability and awareness of many thought resources, such as critical websites and personal blogs, creates new opportunities and difficult situations as people actively use information technology and can find and understand the opinions of others. The sudden increase in activity in the field of thought analysis and mood analysis is at least partly new, dealing with the thinking, mood, and subjective computational processing of textual content. Rescue as a first important subject, he deals primarily with thoughts.

This review describes methods and techniques that promise to search for data in thought-oriented systems. The authors focus on how they try to solve new problems when using emotion-based applications by comparing something that may already exist in traditional fact-based analysis. It includes data on the generalization of the

evaluation text and a wider range of confidentiality, business, and financial implications. This will facilitate the development of event-based information access services. Available resource control datasets and evaluation campaigns will also be discussed to facilitate future work.

Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis (Wilson, Wiebe, & Hoffmann, 2005)

This work presents a new approach to tone level analysis. This first removes the ambiguity of the polarity of the expression, which determines the neutrality or polarity of the next expression. With this method, the device can automatically select the high polarity of the emote sub-environment, resulting in much better results than the original.

This paper introduces a new method for expressing tonality estimation at the expression level. This first determines whether the expression is unbiased or polar and removes the polarity of this polarity expression. In this way, the author mechanically senses the polarity of the context of a huge set of mood expressions and can achieve much better results than the original expression.

2.2 Related Work

Twitter Sentiment Analysis on Demonetization tweets in India Using R language (Arun, Srinagesh, & Ramesh, 2017)

This article analyzes the sentiment of Twitter users, including Democrats, Indians, and the latest news from the country (India) around the world, sharing their opinions on Twitter. The results of the mood analysis were obtained using pie charts. The pie was negative: 38%, neutral: 39%, positive 23%. A positive word cloud sentiment analysis was also conducted to determine how people found the country useful.

Analyzing stock market movements using twitter sentiment analysis (Rao & Srivastava, 2012)

In this article, researchers examine the complex relationship between the tweet literature (e.g., optimism, volume, consent, etc.) and financial market instruments (volatility, volume, stock price, etc.). 4 million tweets from June 2010 to July 2011 against DJIA, NASDAQ-100 and 13 other high-tech companies.

A system for real-time Twitter sentiment analysis of 2012 U.S. presidential election cycle (Hao Wang,2012)

This document describes a real-time analysis of public opinion about 2012 U.S. presidential candidates, presented on Twitter on the microblogging page. The conclusion is that public opinion is changing with new political events or news.

Sentiment analysis and Twitter: a game proposal (Furini & Montangero, 2018)

Sentiment analysis is a difficult task for the computer, but it is amazingly simple for people, so in this article, the author has changed the sentiment analysis step to a game. In fact, they looked at the game in a targeted way and suggested a game where users had to differentiate the polarity and tone of their tweets. As a result, it was stressed that those participants who had a valid way of playing in measuring people's moods liked the game.

Insight on Insiders' threat

We are convinced that in the category of confidential information threats, many confusing organizations do not even have the policies and controls to combat them. Of further concern is the reluctance to approve the insider event, rather to shoot the attackers and protect their reputation. Our understanding is seen as a human error that needs to be addressed at the highest level of internal classification. "A human error is a human act that exceeds certain administrative limits set by the operating system." UIM categories are defined by their outcome or purpose: -

- Unintentional human errors can be caused by a lack of organized knowledge and work skills. This error can inadvertently become another type (intentional or malicious).
- Intentional human error is caused by a user who is aware of the dangerous behavior but will act accordingly or exploit the devices. Improper behavior does not necessarily cause sudden harm to an organization, but it can still violate applicable laws and confidentiality.
- Harmful human error is intentional and the most serious error that will have fatal consequences.

It is easier to blame people for computer events than it is to program computer programs and systems. In fact, you are responsible for designing the system, without taking human factors into account. In many cases, the user who creates the security policy and those who want to apply it have different opinions about the security policy. It is especially important to understand that users often show their own bias when making decisions. This group simplifies the understanding that user training is feasible. Here are some examples: -

- Using public Wi-Fi can lead to unintentional errors in accessing sensitive accounts

and ignoring risks. Or, while working, employees can access dangerous sites related to social networks.

- If a user wants to write a password on a sticker and places it near a computer or desktop drawer and no one uses it, an intentional error can occur.
- Theft (leakage) of employee confidential information can lead to malicious errors.

As mentioned above, user errors can vary depending on the UIM category. For example, users do not allow links and do not download files as unapproved email attachments. If the new employee is unfamiliar with social engineering tactics, you can click on this link (by mistake). On this link, the conversion rate of this employee decreases as they study. Otherwise, the employee's behavior is intentional. Similarly, loot or traps can be used to learn about normal or unusual user behavior. Some companies implement programs that simulate real-life situations, such as fishing practices. We recommend that employees be transparently informed about the use of fishing simulators and other training programs. The goal is not to increase stress in the workload, but to increase the culture of cybersecurity awareness.

(1). To define confidentiality requirements, organizations shall describe their details and location. Users must distinguish between treating public data as confidential or restricted. Data breaches can occur on a user's computer on an external server or storage that is transmitted over an open or closed network. If you change the classification of the data or the status of the user, you must update the user's access to sensitive data. If you see this internal data security requirement as a human error or anomaly, you can help develop authentication policies for anyone who has access to confidential data. For example, use time-based authentication (JIT). The JIT helps to prevent the maintenance of administrative rights. This reduces the risk of administrator credentials being stolen and you will not be able to access administrator data for several hours if you do not need confidential data.

(2). Integrity is a system requirement. The data can be changed by the user for storage or through a closed or open network interface server. If you feel that changing a user policy is incorrect, you can optimally handle not only confidentiality but also consistency. Therefore, you need to examine the impact on user access and system integrity.

(3). Accessibility is also a system requirement. By interconnecting system components, users can influence other parts of the system by influencing the availability of individual systems. If the point of failure is not specified in the system design, user errors can easily occur that intentionally or mistakenly render the system unusable.

2.3 Limitations of prior art

Computer attacks are as common as the Internet itself. Every year, industry reports, the media, and scientific material highlight this growing prevalence, covering both the number and type of cyber-attacks and crimes.

Although there are bundles of research going on currently, on the behavioral perception towards cyber-attack on user, but it has been incredibly challenging for researchers and even scientists to conduct deep study on this and map the users' feelings with possible causation towards future challenges and solution towards the new ways hackers come up with to exploit the system and finding what new digital signatures are required, as cyber security is very unpredictive domain in nature with multitude of new ways being discovered every single day by hackers , the software technology of which is unprepared for to tackle.

Also there have been negligible research regarding the proper and quick resolution by the service provider at the time of such identity theft. So, this study is new in its own way. This study has tried to find out the most common and uncommon malwares and hacker exploitation, known or unbeknownst to the user, through which the most common loopholes in the system can be best targeted towards which have possibly caused the greatest number of thefts and breaches in the user's system, so that organization can pinpoint the serious problem that is responsible for maximum or severe impact to the user's tokens, privileges, identity, or data.

This was all about the introduction part. Next chapters will discuss upon research methodology and approaches, data analysis, limitation, conclusion, and recommendations.

Chapter 3
Research Methodology and Approaches

3.0 Introduction

This section details the methodology and approach of the research, including what the elements of the research are and how the transition to this Twitter sentiment analysis research was done using R.

This section describes the types of sentiment grading methods, the tools used for sentiment analysis, the Twitter sentiment analysis, and finally the Twitter sentiment analysis using R.

3.1 Sentiment Classification Techniques

The most common application of sentiment analysis is the classification of text by class. The classification of sentiment by data set and ratio is a matter of binary (positive or negative) or multiple classes (three or more categories).

The first step in classifying text and keys is preprocessing. The data is used in many ways to reduce text noise, reduce size, and improve sorting efficiency. The most popular techniques are:

- Remove numbers
- Stemming
- Part of speech tagging
- Remove punctuation
- Lowercase
- Remove stop words

There are three ways to rank a sentiment, but it can be said that there are three ways to rank a sentiment. Machine learning, vocabulary (Lexicon-based), hybrid. This is explained in the following order.

3.1.1 Machine Learning

This method creates a classifier that you can use to access cozy(sentimental) texts with access and various machine learning features. Today's deep learning methods are popular because they fit the idea of learning data. Automated systems use machine learning algorithms that learn to predict mood based on previous observations. This AI approach requires a set of data samples (such as analysis data) along with the appropriate tags. This is

called training data. During the learning process, the model converts the text data into vectors (a series of numbers containing coded information, basically what the machine understands) and converts each vector into predefined tags ("Positive", "Negative", "Neutral"). With the right amount of relevant data, the automated system can start its own predictions to classify the invisible data. By increasing the label examples, you can easily increase the accuracy of these models.

3.1.2 Lexicon Based

This method uses many polarity index annotations to determine the overall content rating. The biggest advantage of the method is that no data will be needed for learning, but the weakest factor is that many words and phrases are not included in the synthesis vocabulary. This approach creates a template for each tag using a manually created set of rules. Lexicon is responsible for solving the problem of classifying the atmosphere in the system. The dictionary is a list of positive terms (e.g., good, beautiful, useful, and interesting, etc.) and negative terms (e.g., bad). Ugly, ugly, frustrated, etc.). When loading text, the model counts the number of positive and negative words and assigns the appropriate key. If an expression contains positive and non-negative words, mark it as positive. However, this method has its limitations. It does not recognize words that are not on the vocabulary list, and it is hard to recognize the irony, the satire, that separates words from passages based on context. For example, "Excellent customer service? Welcome!" It can be positively wrong. Finally, rule-based systems are difficult to expand or improve because adding new words to the dictionary can affect previous results.

3.1.3 Hybrid

The combination of machine learning and vocabulary-based sentiment analysis methods is called a hybrid. It is not generally available, but you can get more consistent results from the above. Hybrid systems combine a rule-based approach with machine learning. First, the model learns to define the sentiment based on a series of examples. To improve accuracy, compare the results below with the dictionary. The goal is to achieve the best results without individual access restrictions.

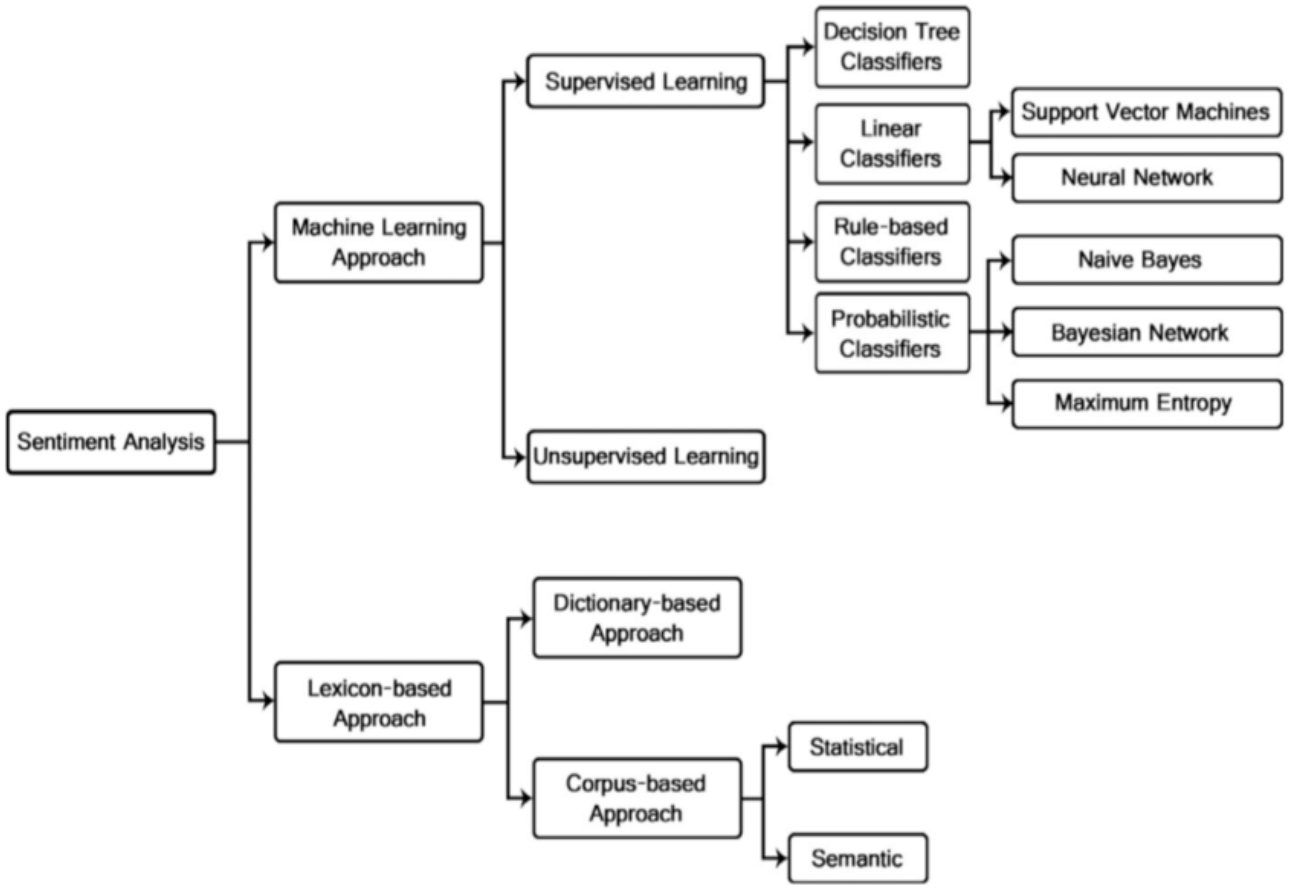

Figure 6: Sentiment analysis Techniques

Source: *https://www.kdnuggets.com/2018/03/5-things-sentiment-analysis-classification.html*

In this study, we used a lexicon-based approach. This is because the corpus is formed, and its sentences and words are analyzed correctly.

3.2 Categories of Sentiment Analysis

Sentiment analysis models focus not only on polarity (positive, negative, neutral) but also on emotions and feelings (anger, happiness, sadness, etc.) and even intentions (e.g., if interested).

It presents the most popular types of sentiment analysis.

3.2.1 Fine-Grained Sentiment Analysis

The degree of polarity is beneficial for most businesses and organizations. They may consider certain categories to extend their polarity. The categories are as follows:

• extremely negative
• negative

• Neutral
• positive
• incredibly positive

This is usually called a detailed sentiment analysis,

It serves to interpret the five-star rating as a critique. Example:

• Incredibly positive = 5 stars.
• Extremely negative = 1 star.

3.2.2 Emotion Detection

The purpose of this type of sentiment analysis is to explore emotions such as happiness, frustration, anger, sadness, and so on. Many emotion detection systems use dictionaries (that is, a list of words and emotions they convey) or complex algorithms to get used to the device.

One of the disadvantages of using the Lexicon is that people express their emotions in their own way. Words that usually express anger, such as horrible murder (such as a product that kills with serious support services), can also be unique (such as a terrifying criminal or murder).

3.2.3 Aspect-based Sentiment Analysis

Fine-grained analysis helps determine the overall polarity of customer feedback, while aspect-based analysis allows for deeper analysis. This will help you understand the specific aspect that people are talking about. Let us say this. You are a cell phone manufacturer and have a customer opinion that "cameras do not work well in artificial lighting". The prospective analysis will allow critics to judge that they have made “negative” comments about “cameras”.

3.2.4 Multilingual Sentiment Analysis

Multilingual analysis of comments can be daunting. This requires a lot of preparation and resources. Most of these resources are available on the Internet (such as sentiment Lexicon) and other resources need to be written (such as translated enclosures and noise detection algorithms) and you need to know how to use them.

3.3 Sentiment Analysis Tools

Using off-the-shelf tools and APIs: - Various Software’s (In Moment, Clara Bridge, etc.) collect notes from a variety of sources and send notifications for real-time reports for text analysis and display of results. Text analysis platforms (Discover Text, IBM Watson Natural

Language Understanding, Google Cloud Natural Language, Microsoft Text Analytics API, etc.) include sentiment analysis in features.

In Moment: - In Moment offers five products that provide a platform to optimize the customer experience. One is Customer Voice, which allows organizations to collect and analyze customer feedback in text, video, and audio formats. The number of data sources is sufficient, consisting of research, social networks, CRM, etc. Developers provide customers with real-time, customizable message boards and several reporting options.

Clara bridge: - Clara Bridge is a customer experience management (CEM) platform. You can extract and analyze textual content from chat platforms, blogs, forums, and review sites. Users can receive notes, recorded names, or interactive voice response (IVR) surveys via e-mail, from employees and operators. The system can convert them to text content. It also offers urgent social media. The system understands the importance and context of each comment, considering the industry and source. The results of the sentiment analysis are displayed on an 11-point scale. Depending on their needs, users can customize their ratings to apply only to the business.

Discover Text: - Discover Text is a complete cloud-based text analysis system for researchers, businesses, and governments. Capterra users note that this solution is ideal for importing / searching, filtering, and analyzing data from a variety of sources, such as Twitter, SurveyMonkey, email and spreadsheets. Sentiment Analysis is several ways to analyze textual content in Discover Text.

IBM Watson Natural Language Understanding: - IBM Watson Natural Language Understanding is a set of advanced text analysis systems. By analyzing the textual content of the service, users can use metadata such as conceptual entities, keywords, categories, and relationships. In addition, it is possible to identify the author's feelings and sentiments, which vary depending on which text the document belongs to and the semantic role of the domain proposal section. IBM Watson Natural Language Understanding supports resolution in 13 languages. Visitors can use IBM Watson Services to order development tools to create their own solutions (Chatbots, etc.).

Microsoft Text Analytics API: - Microsoft Text Analytics API User keyword term name (person, company, location, etc.) Have sentiment You can check that the text was written in 120 supported languages. The sentiment analysis results from API 0 (performance) to 1 (positive) (Sentiment Analysis Score). Currently, the program can enter the sentiment text key in English, Spanish, German, and French. Developers are encouraged to use a one- or two-sentence document to achieve high accuracy, indicating that these should be implemented in the detailed document. See how Microsoft Text Analytics API analyzes the opinion of the movie "Nun": it searches English with 100% confidence and its sentiment is measured as a percentage. The analysis results are also returned in JSON format.

Google Cloud Natural Language API: - The Google Cloud Natural Language API exports

sentiments from mail, text documents, news articles, social media, and blog posts. Its use includes audio files, scanned documents, and other cloud services, as well as the ability to export information from documents in different languages.

Twitter API: - The Twitter API is a set of custom URLs. Use this URL to access many Twitter features, such as tweeting or searching for tweets.

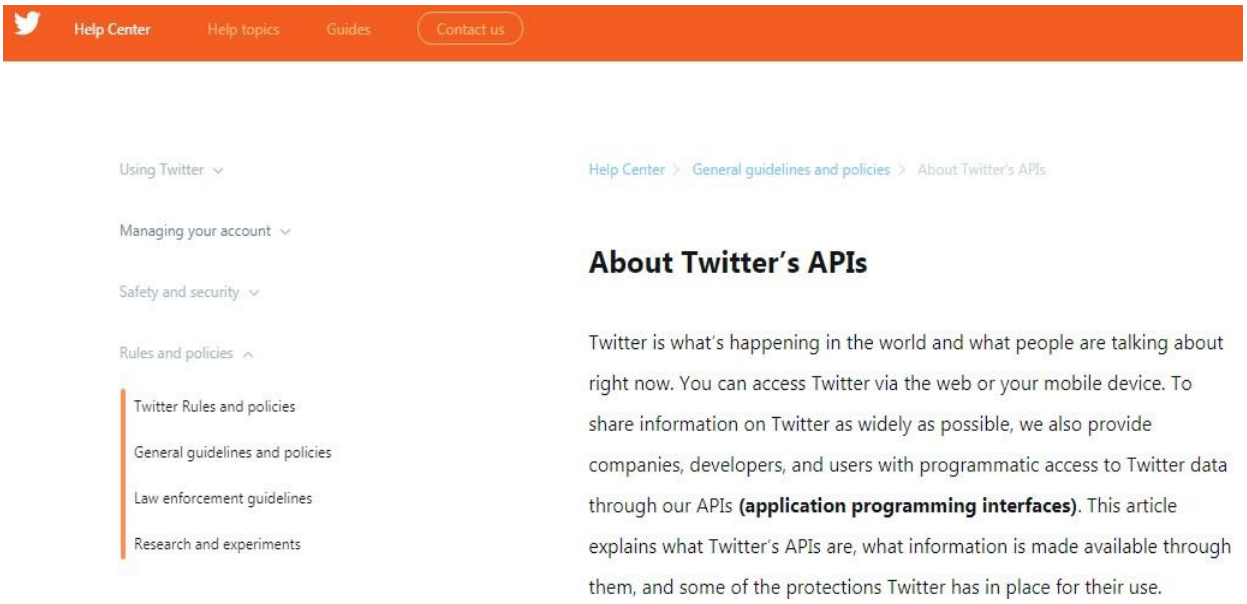

Figure 7: Twitter API Help Center

Source: *https://help.twitter.com/en/rules-and-policies/twitter-api*

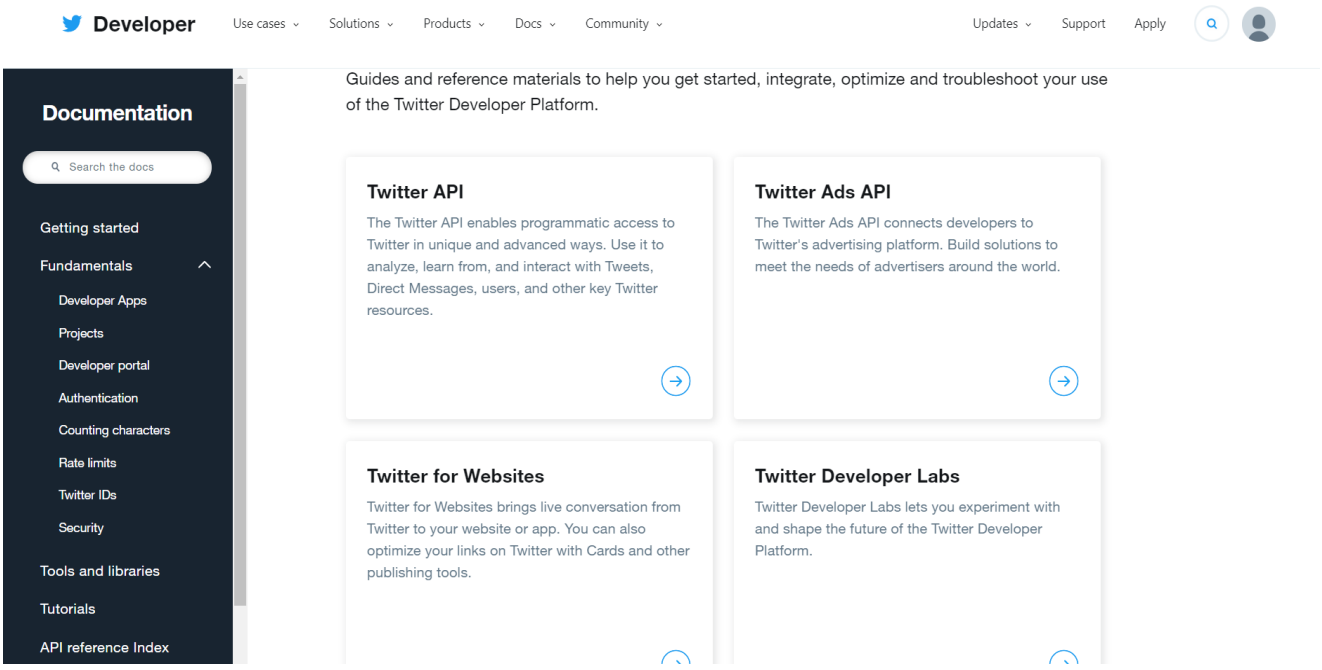

Figure 8: Twitter Developer Interface

Source: *https://developer.twitter.com/en/docs*

You can get the keys and badges you need to write a Twitter app and access the Twitter developers you want in your Twitter developer account. In this survey, we used the Twitter platform and used Twitter API to access tweets.

3.4 Sentiment Analysis Methodology

There are 5 steps to analyzing sentiment data. This is a graph of a methodology that can do the same.

Figure 9: Methodology for sentiment Analysis

Source: *https://www.researchgate.net/figure/Sentiment-Analysis-Methodology_fig1_330880816*

Methodology for Sentiment Analysis

- **Data Collection:** - Consumers typically express their sentiments in public forums, such as blogs, forums, and product brochures, or in personal magazines on social networking sites, such as Facebook or Twitter. Thoughts and feelings are expressed through vocabulary, writing, short form, and slang, and various data occur. It is almost impossible to manually analyze availability of sentiment data. Therefore, special programming languages such as "R" are used for data processing and analysis.
- **Text Preparation:** - Preparing the text is no different from filtering the exported data before analysis. This includes finding and removing non-textual content and removing content without studying the data.
- **Sentiment Detection:** - In this section, we check the subjectivity of each withdrawal proposal and thought. Subjective expression suggestions are saved, and objective expression suggestions are rejected. Sentiment analysis is performed at different levels using general calculation methods such as unigram, lemmas, and negative methods.
- **Sentiment Classification:** - Sentiments can usually be divided into two groups: positive and negative. Sentiment Analysis Methodology All subjective suggestions defined in this section are classified of positive, negative, good, bad, like, and dislike.
- **Presentation of Output:** - The main idea of sentiment analysis is to turn unstructured text into important information. When the analysis is complete, the text results are displayed in charts, such as pie charts, histogram bar charts, and so on. At present, sentiment analysis is an important task for any supplier of a product or service. So, let us start using the "R" language!

3.5 Twitter Sentiment Analysis using R

Here are the steps used in the project to achieve the research goals.

1. After acquiring Tweet using the Twitter API, the authentication process was done using the RStudio platform.
2. Tweet pretreatment and cleaning.
3. Analysis of tweet sounds.
4. Observing through the graphs which cyber-attack or theft causes the most severe reaction in terms of emotion exhibition among users. When the user is enraged or devastated over lost data, that is when the severity meter is towards higher side.
5. Devise the possible resolution strategy at the time of attack, resolution of aftermath and targets to be pinpointed at to avoid any future cyber crisis by using text mining.

3.5.1 Tools and Packages Used

For this research paper or study, RStudio GUI and several other packages been used. These packages are: -

- **twitteR: -** This package provides a Twitter Web API interface**.**
- **ROAuth: -** This package provides an OAuth 1.0 interface and allows users to authenticate through OAuth to a server of their choice.
- **plyr**: - This package is a set of tools that solves common problems. You need to break down the big problem into manageable parts, take care of each part, and clean all the parts.
- **stringr**: - The stringr is a simple shell set that makes the R string function more consistent, simpler, and easier to use. This is done consistently with the name and argument of the function (and positions) and matching the output

structure of each function to the input structure of another function, while all functions function as NA characters and zero characters.

- **ggplot2**: - Graphic application in R. It has the best graphics and graphics grid in it. Mapping and common axes are automatically addressed, and you can create step-by-step charts from multiple data sources.
- **RColorBrewer**: - Provides a palette for designing appropriate color maps based on package variables.
- **tm**: - This is the basis of the R. text mining application.
- **wordcloud**: - This package will help you create a beautiful word cloud in text mining.
- **Syuzhet**: - This package extracts sentiment and sentiment-derived plot arcs from text using a variety of sentiment dictionaries conveniently packaged for consumption by R users.
- **sentimentr**: - This package is designed to quickly calculate text polarity sentiment at the sentence level and optionally aggregate by rows or grouping variable(s).
- **qdap**: - (Quantitative Discourse Analysis Package) is an R packet designed to support quantitative speech analysis. This package serves as a bridge between high-quality dialogue scenarios and statistical analysis and visualization.

Chapter 4
Data Analysis

4.0 Introduction

This section describes the actual analysis of Tweets using R Studio. This section describes the various packages and directories that can be used to activate the various required features that are installed. This section describes practical programming using R. The results obtained after executing the command in studio R are analyzed later.

This section describes each step of the data analysis. Steps: data collection, data processing, text mining, sentiment analysis, interpretation of results.

4.1 Data Collection

- ✓ **Establish twitter API connection through R**

Use the OAuthFactory function and Twitter credentials to connect to the Twitter API.

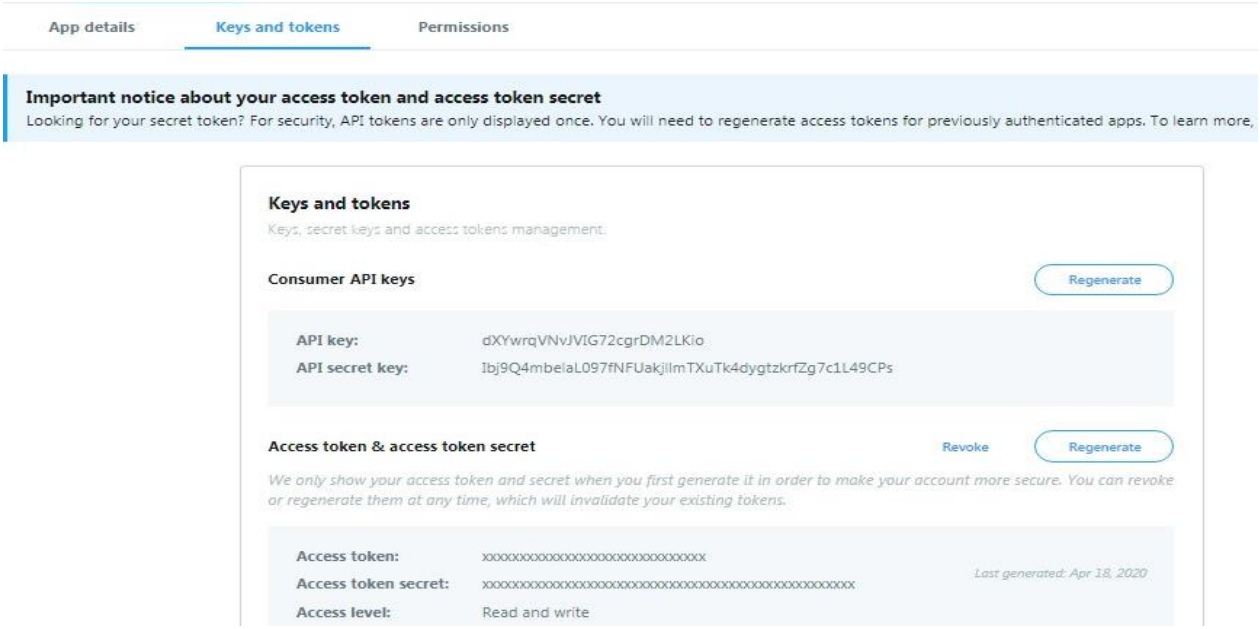

Figure 10: Twitter API Keys and Tokens (Personal Twitter Account)

✓ **Extract tweets from Twitter API**

To export a tweet, use the search_tweets function to convert the data to a data frame using the as.data.frame function.

```
library(sentimentr)
library(qdap)
oauth_endpoint(authorize = 'https://api.twitter.com/oauth', access = 'https://api.twitter.com')
reqURL = 'https://api.twitter.com/oauth/request_token'
accessURL = 'https://api.twitter.com/oauth/access_token'
authURL = 'https://api.twitter.com/oauth/authorize'
consumer_APIkey = "m16JxgvgkqDLETUSwB6VULydl"
consumer_APIsecret = "0ydx8AdNzmleemZt3eJPGK0wnew0bdSuSXCNnxX7xbA11ahFOQ"
access_APItoken = "968416291785773056-Yh8kcPQkKjNp9RmiMxXiLkOoyFIoemj"
access_APIsecret = "JXVJ8HN7vmn2N39zy8ox3xaAPpwi861dNoawyg5TRNbOH"
Cred <-OAuthFactory$new(consumerKey = consumer_APIkey, consumerSecret = consumer_APIsecret,
Cred$handshake(cainfo = system.file('CurlSSL', 'cacert.pem', package = 'RCurl'))
save(Cred, file = 'twitter_authentication.Rdata')
load('twitter_authentication.Rdata')
setup_twitter_oauth(consumer_APIkey, consumer_APIsecret, access_APItoken, access_APIsecret)
tweets<- search_tweets("hijacked account", n = 3000, type = "recent", lang = "en", incl
tweets.length = length(tweets)
tweets.length
```

4.2 Data Preprocessing

Data Cleaning

Following steps are executed to clean the tweets text.

- The first step in this direction is to remove the URL from punctuation, numbers, etc. After removing as Removing DURL can be difficult.
- Delete the line between "<" and ">". This is used to remove emoticons and other encoded text.
- Remove retweet entities like @RT etc.
- Remove quotes and apostrophes e.g., ‘Scammer’, ‘Phishing’, and so on.
- Delete punctuation marks. This will delete the basic English sentences.
- Delete single alphabet characters, such as "s" and "a".
- Remove excess space.

```
tweets.df = as.data.frame(tweets)
tweets.df$OriginalText = tweets.df$text
head(tweets.df)
tweets.df$text = gsub("http[^[:space:]]*", "",tweets.df$text)
tweets.df$text = gsub("(RT|via)((?:\\b\\W*@\\w+)+)"," ",tweets.df$text)
tweets.df$text = gsub("'s|'s|[...]", "", tweets.df$text)
tweets.df$text = gsub("@\\w+", " ", tweets.df$text)
tweets.df$text = gsub("[[:punct:]]", " ", tweets.df$text)
tweets.df$text = gsub(" *\\b[[:alpha:]]{1}\\b *", "", tweets.df$text)
tweets.df$text = gsub("[ \t]{2,}", " ", tweets.df$text)
tweets.df$text = gsub("^\\s+|\\s+$", "", tweets.df$text)
str(tweets.df)
```

Corpus Creation

A corpus is a set of textual documents to which word processing or natural language processing procedures are applied to draw conclusions. The tm R package can create a corpus of files or vectors.

```
tweets.text = tweets.df$text
myCorpus<- Corpus(VectorSource(tweets.text))
myCorpus = tm_map(myCorpus, removePunctuation)
removeSingle <- function(x) gsub(" . ", " ", x)
myCorpus = tm_map(myCorpus, removeWords, c(stopwords("english"),c('The', 'Rs'
myCorpus = tm_map(myCorpus, stripWhitespace)
myCorpusCopy<- myCorpus
myCorpus<-tm_map(myCorpus, stemDocument)
stemCompletion2 <- function(x,dictionary) {
  x <- unlist(strsplit(as.character(x)," "))
  x <- x[x !=""]
  x <- stemCompletion(x, dictionary = dictionary)
  x <- paste(x, sep="", collapse=" ")
  PlainTextDocument(stripWhitespace(x))
}
```

4.3 Sentiment Analysis

Two sets of tweets are tested for analysis: -

The first set of tweets that uses "Hijacked Account" as a search bar (let us denote it with M_Tweets).

The second set of tweets that uses the "Malware" as a search bar (let us denote it with T_Tweets).

Sentiment Analysis

```
tweets <- iconv(tweets, from="UTF-8", to="ASCII", sub="")
tweets <-gsub("(RT|via)((?:\\b\\w*@\\w+)+)","",tweets)
tweets <-gsub("@\\w+","",tweets)

ew_sentiment<-get_nrc_sentiment(tweets)
sentimentscores<-data.frame(colSums(ew_sentiment[,]))
names(sentimentscores) <- "Score"
sentimentscores <- cbind("sentiment"=rownames(sentimentscores),sentimentscores)
rownames(sentimentscores) <- NULL
ggplot(data=sentimentscores,aes(x=sentiment,y=Score))+
  geom_bar(aes(fill=sentiment),stat = "identity")+
  theme(legend.position="none")+
  xlab("Sentiments")+ylab("Scores")+
  ggtitle("Total sentiment based on scores")+
  theme_minimal()
```

By analyzing the sentiments of the two tweet sets, you can get a graph as follows:

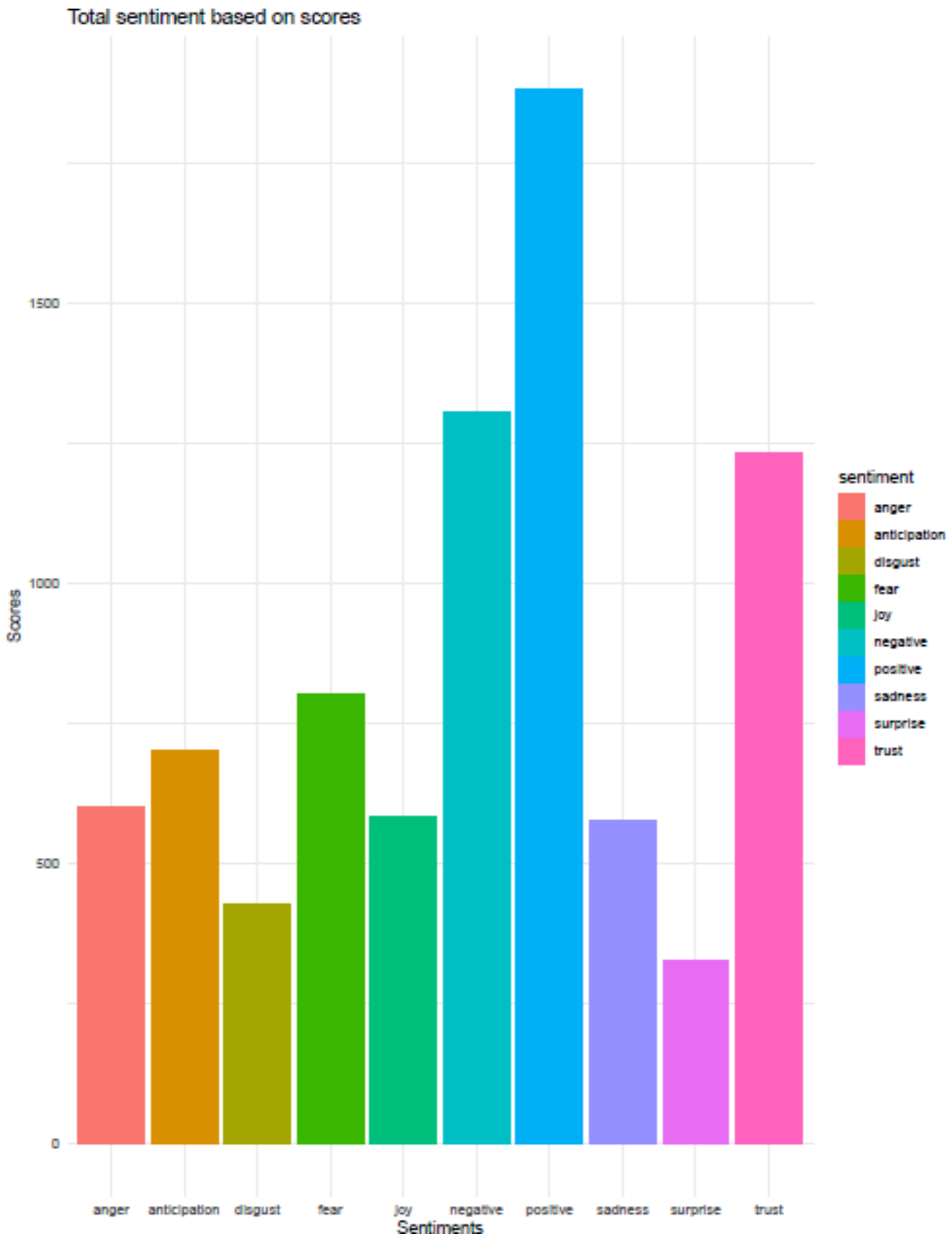

Figure 11: Sentiment Analysis of M_Tweets (own analysis)

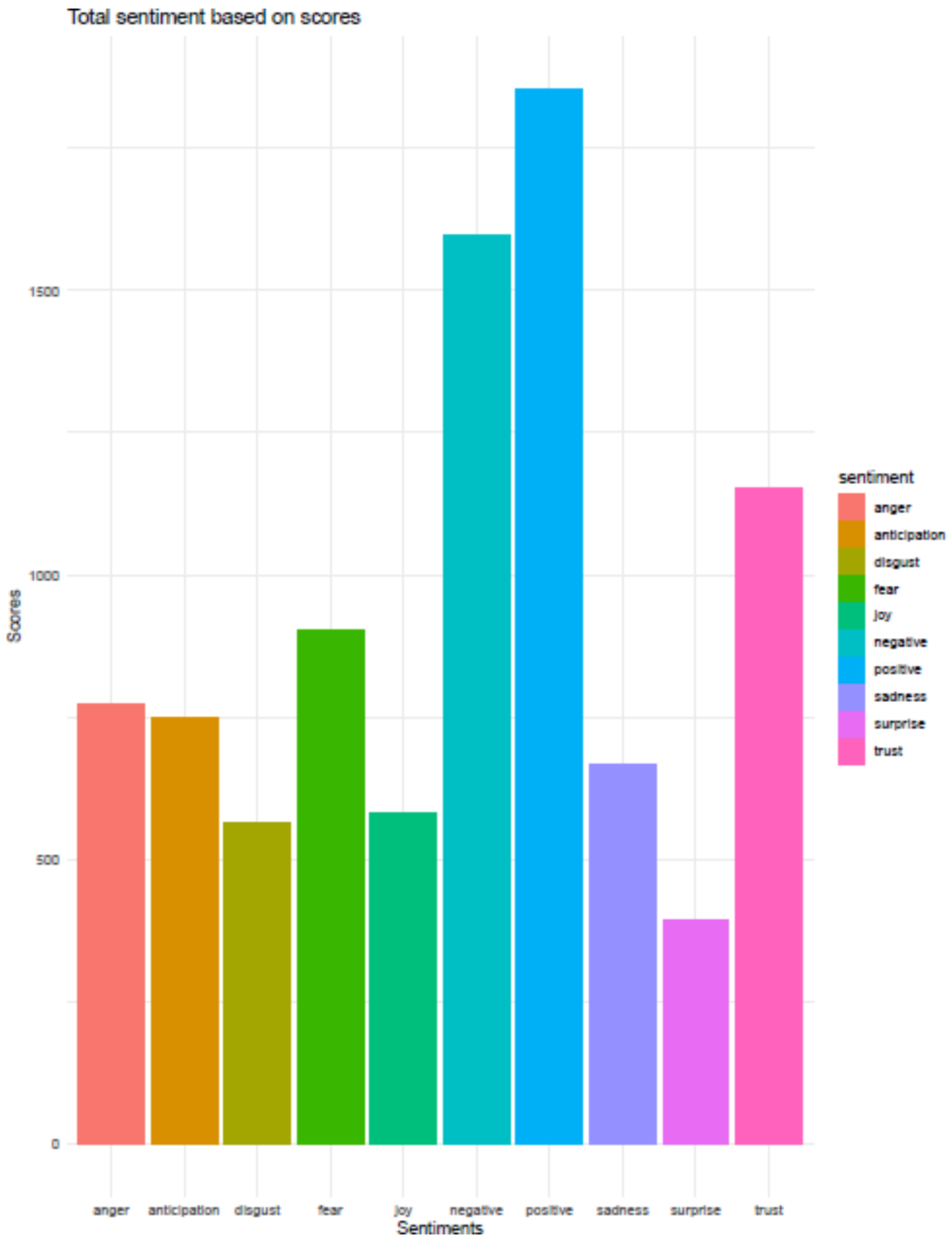

Figure 12: Sentiment Analysis of T_Tweets (own analysis)

4.4 Text Mining

Let us assume that M_Tweets contains preprocessed clean tweets.

Let us assume that T_Tweets contains preprocessed clean tweets.

```
myCorpus <- lapply(myCorpus, stemCompletion2, dictionary=myCorpusCopy)
myCorpusAfterStemComplete = myCorpus
myCorpus <- Corpus(VectorSource(myCorpus))
tdm <- TermDocumentMatrix(myCorpus, control= list(wordLengths= c(1, Inf)))
freq.terms <- findFreqTerms(tdm, lowfreq = 50)
term.freq <- rowSums(as.matrix(tdm))
term.freq <- subset(term.freq, term.freq > 100)
df <- data.frame(term = names(term.freq), freq= term.freq)
ggplot(df, aes(reorder(term, freq),freq)) + theme_bw() + geom_bar(stat = "identity'
```

Text mining can be performed after the corpus is created with the pre-processed data. Let us look at the functions used here.

lapply: - This function returns a list of the length equivalent to mycorpus. All elements are the result of applying stemcompletion2 to the mycorpus element.

findFreqTerms: - You can find common expressions in the document expression or expression-document matrix with the help of this function.

Thus, by executing the above code, the graph contains a graph containing general expressions with a frequency above 100.

This graph is displayed after executing the above code.

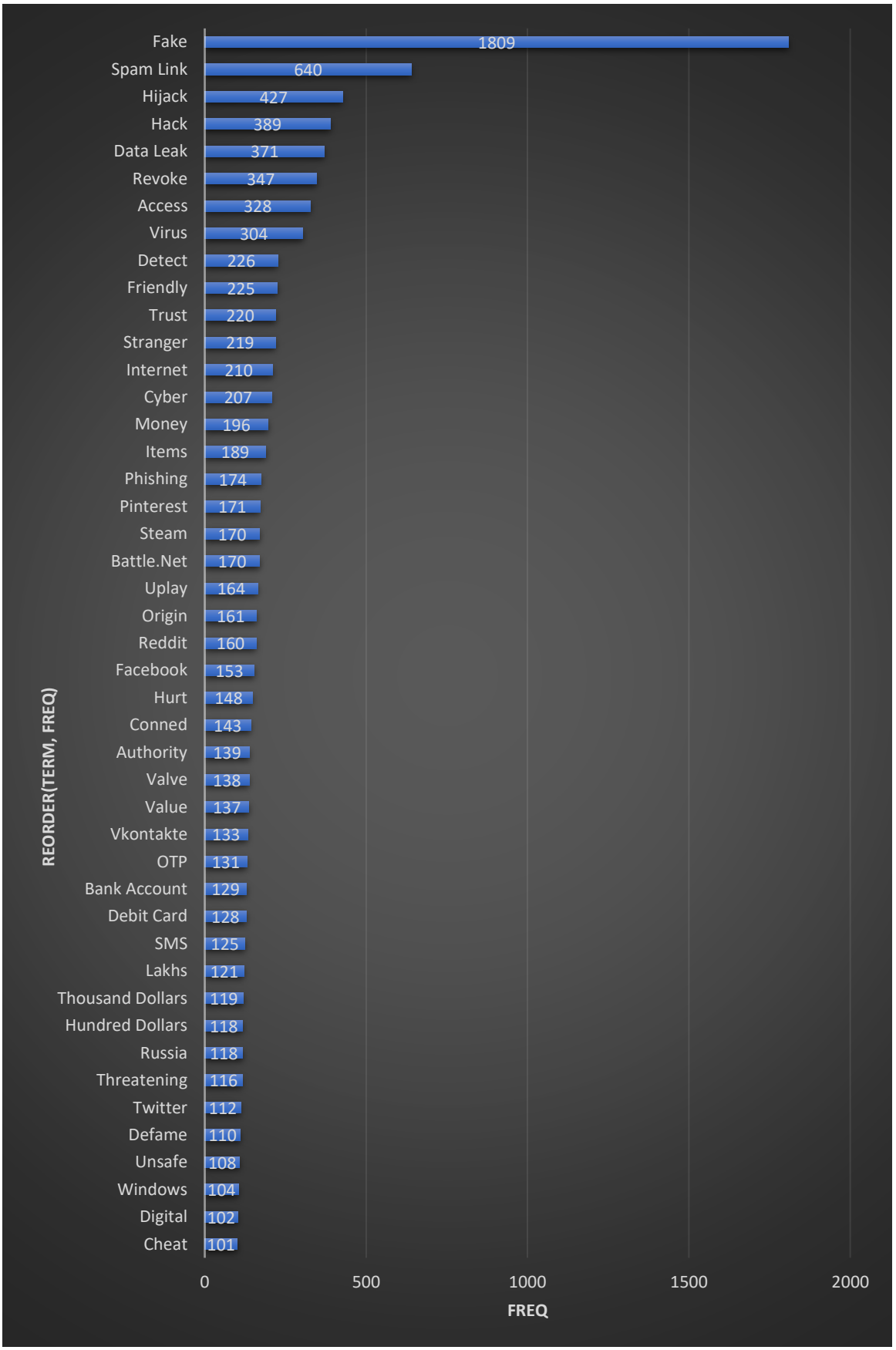

Figure 13: Text Mining of M_Tweets (extracting words whose frequency is greater than 100)

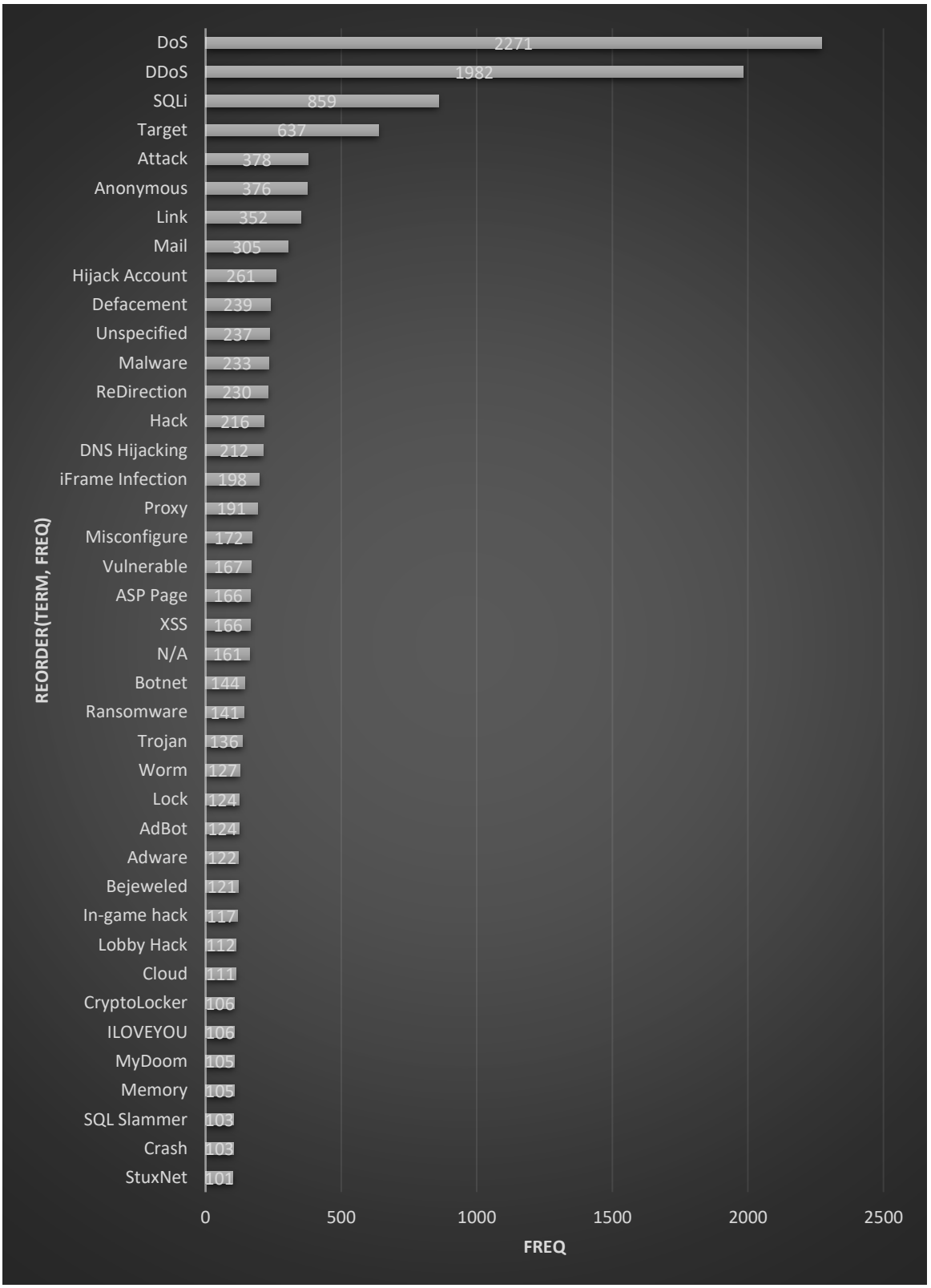

Figure 14: Text Mining of T_Tweets (extracting words whose frequency is greater than 100)

4.5 Result Interpretation

Sentiment Analysis Graph

Two sets of tweets are tested for analysis: -

The first set of tweets that uses "Hijacked Account" as a search keyword (let us denote it with M_Tweets).

The second set of tweets that uses the "Malware" as a search keyword (let us denote it with T_Tweets).

Table 1: Own Analysis (Comparative/Relative Sentiments)

Sentiments	M_Tweets	T_Tweets
Anger	Lower	Higher
Anticipation	Slightly Lower	Slightly Higher
Disgust	Lower	Higher
Fear	Slightly Lower	Slightly Higher
Joy	Almost same	Almost same
Negative	Lower	Higher
Positive	Almost same	Almost same
Sadness	Slightly Lower	Slightly Higher
Surprise	Slightly Lower	Slightly Higher
Trust	Slightly Higher	Slightly Lower

Higher / Lower: When the sentiment difference exceeds 300 score.

Slightly Higher / Lower: For a difference in sentiment score between 100-200 points.

Almost Same: If the sentiment difference is less than 50 points.

So, let us talk about T_Tweets (using the “malware” search keyword) and find out that people are in a positive mood compared to M_Tweets (using the “hijacked account” search bar) because feelings like expectations, disgust, sadness, and the surprise is at the upper side for M_Tweets.

Now you can see how people react to cyber-attacks. If so, you can determine if something really caused that feeling. Thus, by analyzing the text, we identify high-frequency words that correspond to two sets of tweet data (M_Tweets and T_Tweets).

Text Mining Graph

Two sets of tweets are tested for analysis: -

The first set of tweets that uses "Hijacked Account" as a search keyword (let us denote it with M_Tweets).

The second set of tweets that uses the "Malware" as a search keyword (let us denote it with T_Tweets).

***Table 2**: Some relevant words whose frequency was even greater than 160 (Own analysis from Figure 4 & Figure 5)*

M_Tweets	T_Tweets
Fake	DoS
Spam	DDoS
Hijack	SQLi
Hack	Target
Data Leak	attack
Revoke	Anonymous
Access	Link
Virus	Mail

Do the organizations respond to the query posted by users and provide resolution to their issues?

So now, we are done with the Sentiment Analysis and Text Mining part. Now, let us see if the respective organizations servicing the user with pre-configured security software installed in the system can provide resolution to the users for their issues. The higher the value in pie chart, the better the organization's engagement and customer satisfaction.

Below are pie chart curves of the two sets of tweets (M_Tweets (Hijacked Account) and T_Tweets (Malware)).

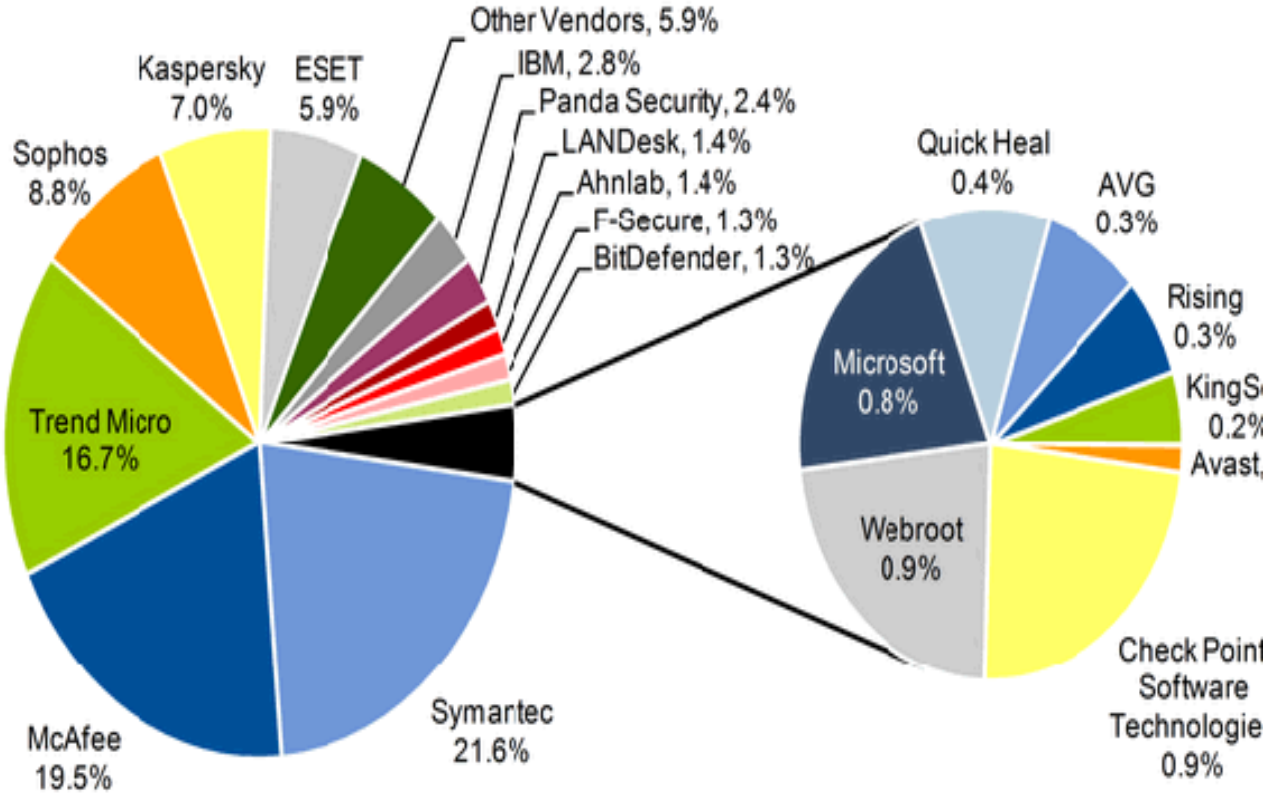

Figure 15: Cyber Security Companies Trustworthiness and Customer Engagement Pie Chart

Source: *https://www.usatoday.com/story/cybertruth/2013/10/07/video-series-depicts-cybercrime-in-2020/2938513/*

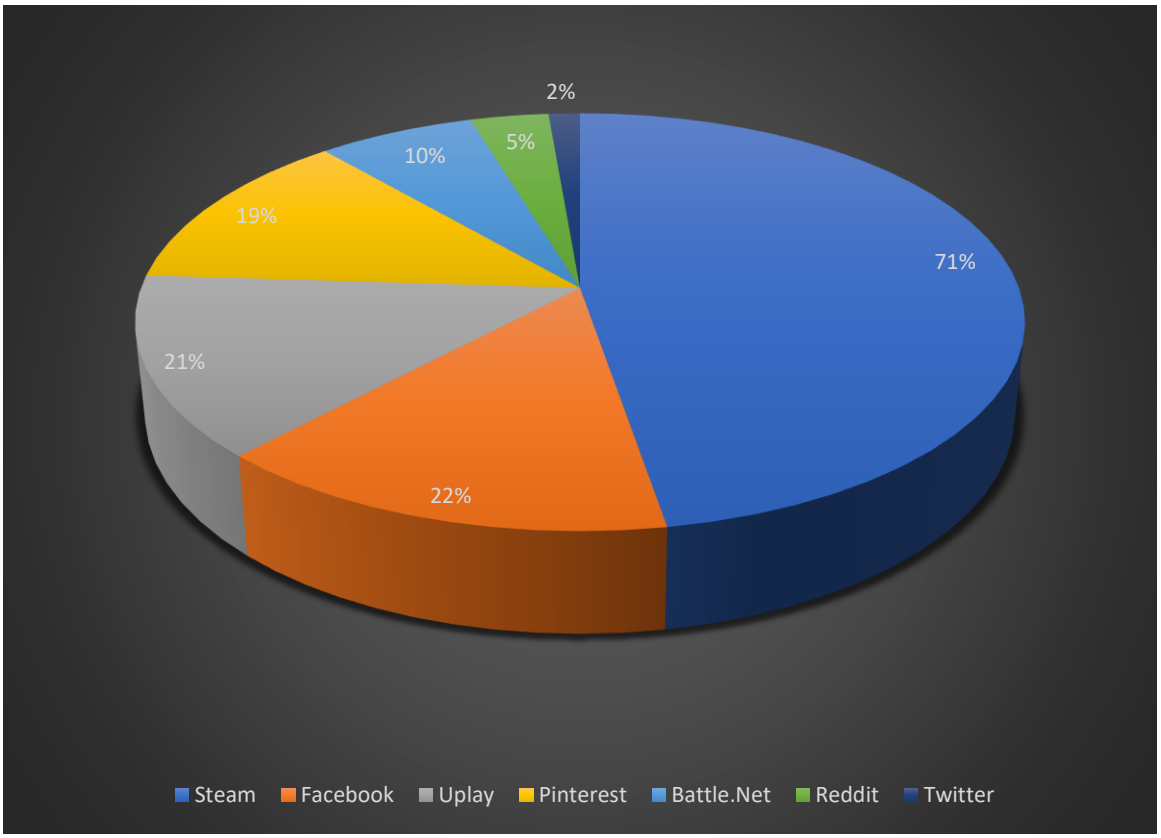

Figure 16: Distribution of social media users affected through account hijack on a particular platform

Source: *https://www.yourlifeupdated.net/windows/microsoft-consiglia-agli-utenti-di-non-usare-lantivirus-microsoft-security-essentials/*

Through graphs, it can be observed that in case of Hijacked account, hackers or cyber hackers are most interested in hijacking user's Steam (a gaming platform). And the other graph shows, the customer satisfaction and engagement from the companies that provide virus protection services to the users in case any attack or data breach happens in the user's workstation or System. That graph shows that Symantec that made the Norton Anti-Virus Engine is the most used and trustworthy platform as well as their organization's engagement and response is maximum in terms of quick resolution to user's response.

This section deals with the analysis of the data and the interpretation of the results. The following sections describe limitations, conclusions, and suggestions.

Chapter 5
Conclusion, Recommendations & Limitations

5.0 Introduction

So far, the analysis part has been carried out with some results achieved in the previous stage. This section describes conclusions, suggestions, and limitations.

5.1 Conclusion

Aspects of cybersecurity behavior are becoming an important area of research. Because of human behavior and the unpredictable nature of behavior, people are important factors and tools for ensuring a level of cybersecurity. Discuss these theories - Stresses the importance of social factors, behavior, circumstances, prejudices, perceptions, coercion, intentions, behavior, decisions about alternative sanctions. Understanding cybercrime. This theory has some limitations but can be used together to reinforce the behavioral model. You need to understand and shape the behavior and intentions of users and criminals. Improving this area will prevent unforeseen accidents. It is impossible to 100% protect the system, but it is impossible to provide maximum security without taking human factors into account. The slogan quoted by President Ronald Reagan is "trust only", which refers to cybersecurity. There is a degree of trust in cyberspace to deal with it, but it requires constant monitoring. Employees need to be aware of the risks and differentiate between desirable and undesirable behaviors. However, some employees still do not use neutralization techniques. Computer literacy training needs to be made unique because employees may have different levels of power or access and responsibility. They also have their own security prejudices. Universal training programs do not work. Employees need to be given some degree of confidence, but they need to be taught cyberspace skills and awareness and checked for compliance. For multiple trainings, this method can only be the solution. A comprehensive conceptual framework that combines cybersecurity behavior, human factor modeling, and simulation is proposed. For the model to work as intended, companies need to be included in a survey. There are times when it is wrong to use a comfortable model without customization. Coordinating cyberspace behavior and technical security should be common to all organizations.

5.2 Recommendation

As cyber threats and cyber-attacks continue to permeate the Internet, it is important as a community to have a good understanding of this issue and how it affects our lives. How is this work and research progressing towards this goal? Citizens are aware of the dangers, and not just their faces, but also how they may be affected by a cyber-attack. Because attacks are often overlooked in research and practice, the focus has been on the social and psychological effects of attacks. But this is especially important. Factors that understand the broader aspects of the impact of an attack. To justify our work, we examined two well-known cyber-attacks, so we examined them in a broader context. This study is expected to encourage others to continue exploring the interactions between cybersecurity and cognitive factors in this area. We invest heavily in the information security of corporate organizations around the world. While these investments are critical, knowledge of how to ensure appropriate information security awareness and behavior Organizing is still difficult. The results of this dissertation can help researchers develop future research strategies to better understand how to change information security behavior. Those responsible for information security also have a real impact. The results can support them in the decision-making process by understanding how to address general information security threats, especially behavioral threats.

5.3 Limitations

There are several limitations to this study.

- Only 3,000 tweets have been extracted via the Twitter API using search_tweets (), and results may be limited.
- All tweets that contain keywords are included. There are no restrictions on where you can get tweets from a particular site. Because we analyzed people's attitudes toward system theft or cyber-attacks, it is a good idea to use the method of extracting Tweets from certain sites.
- In this research, the search string was limited to "Hijack Account", but piracy, robbery, hacking, etc. can also be used.
- This study does not involve multilingual analysis, such as tweets in other languages. Only English tweets were analyzed.

These constraints were intentionally introduced to accommodate all volumes and reduce complexity. These limitations can be overcome if the necessary resources, such as human resources, time, and skills, are available, but current research circumvents all of this to complete this research in a timely manner with the available resources.

REFERENCES

https://www.cyberintelligence.in/top-ten-countries-with-weak-cyber-security/

https://www.nature.com/articles/d41586-020-00758-2

https://www.kaspersky.com/resource-center/definitions/what-is-cyber-security

https://digitalguardian.com/blog/what-cyber-security

https://www.kdnuggets.com/2018/03/5-things-sentiment-analysis-classification.html

https://help.twitter.com/en/rules-and-policies/twitter-api

https://developer.twitter.com/en/docs

https://www.researchgate.net/figure/Sentiment-Analysis-Methodology_fig1_330880816

https://www.usatoday.com/story/cybertruth/2013/10/07/video-series-depicts-cybercrime-in-2020/2938513/

https://www.yourlifeupdated.net/windows/microsoft-consiglia-agli-utenti-di-non-usare-lantivirus-microsoft-security-essentials/

https://www.nap.edu/read/25335/chapter/10#177

BIBLIOGRAPHY

Bibliography

Abbasi, A., and Chen, H. (2008).
Cybergate: A design framework and system for text analysis of computer-mediated communication. MIS Quarterly: Management Information Systems

Agarwal, N., and Bandeli, K.K. (2017).
Blogs, fake news, and information activities. In G. Bertolin (Ed.), Digital Hydra: Security Implications of False Information Online

Ahn, J., Taieb-Maimon, M., Sopan, A., Plaisant, C., and Shneiderman, B. (2011). *Temporal visualization of social network dynamics: Prototypes for nation of neighbors. In International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction*

Al-Khateeb, S., and Agarwal, N. (2016).
Understanding strategic information maneuvers in network media to advance cyber operations: A case study analyzing pro-Russian separatists' cyber information operations in Crimean water crisis. Journal on Baltic Security

Al-Khateeb, S., Hussain, M.N., and Agarwal, N. (2017). Analyzing deviant socio-technical behaviors using social network analysis and cyber forensics-based methodologies. In O. Savas and J. Deng (Eds.), Big Data Analytics in Cybersecurity and IT Management

Allcott, H., and Gentzkow, M. (2017).
Social media and fake news in the 2016 election. Journal of Economic Perspectives

Alter, A.L., and Oppenheimer, D.M. (2009).
Uniting the tribes of fluency to form a meta-cognitive nation. Personality and Social Psychology Review

Altman, N., Carley, K.C., and Reminga, J. (2018).
ORA User's Guide 2018. Technical Report

Alvanaki, F., Michel, S., Ramamritham, K., and Weikum, G. (2012).

See what's enBlogue: Real-time emergent topic identification in social media.

Amarasingam, A. (Ed.). (2011).
The Stewart/Colbert Effect: Essays on the Real Impacts of Fake News.

Anderson, K.E. (2017).
Getting acquainted with social networks and apps: Social media in 2017.

Aral, S., and Van Alstyne, M. (2011).
The diversity-bandwidth trade-off. American Journal of Sociology

Asur, S., and Huberman, B.A. (2010).
Predicting the future with social media.

Babcock, M., Beskow, D., and Carley, K.M. (2018).
Beaten up on Twitter? Exploring fake news and satirical responses during the Black Panther movie event.

Baggili, I., and Breitinger, F. (2015).
Data sources for advancing cyber forensics: What the social world has to offer. In Sociotechnical Behavior Mining: From Data to Decisions?

Bakshy, E., Messing, S., and Adamic, L.A. (2015).
Exposure to ideologically diverse news and opinion on Facebook.

Barger, V.A., and Labrecque, L. (2013).
An integrated marketing communications perspective on social media metrics.

Bean, H. (2011).
No More Secrets: Open-Source Information and the Reshaping of U.S. Intelligence.

Benigni, M., Joseph, K., and Carley, K.M. (2017a).
Mining online communities to inform strategic messaging: Practical methods to identify community-level insights.

Benigni, M., Joseph, K., and Carley, K.M. (2017b).
Online extremism and the communities that sustain it: Detecting the ISIS supporting community on Twitter.