ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE INGENIERÍA DE SISTEMAS

GROOMING: ANALYZING THE BEHAVIOR OF ONLINE ATTACKERS FROM THE PERSPECTIVE OF SOCIAL ENGINEERING

THESIS SUBMITTED AS PART OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF DOCTOR OF PHILOSOPHY IN INFORMATICS

PATRICIO XAVIER ZAMBRANO RODRIGUEZ
patricio.zambrano@epn.edu.ec

SUPERVISOR: PHD. JENNY GABRIELA TORRES OLMEDO
jenny.torres@epn.edu.ec

Quito, enero 2023

ESCUELA POLITÉCNICA NACIONAL

THESIS

For the award of the degree of

DOCTOR OF PHILOSOPHY IN INFORMATICS

Resolution RPC-SO-43-No.501-2014 of the Consejo de Educación Superior

Presented by

PATRICIO XAVIER ZAMBRANO RODRIGUEZ

Thesis supervised by **PhD. Jenny Gabriela Torres Olmedo, Professor of the Escuela Politécnica Nacional**

GROOMING: ANALYZING THE BEHAVIOR OF ONLINE ATTACKERS FROM THE PERSPECTIVE OF SOCIAL ENGINEERING

Oral examination by the following committee:

Luis Enrique Sánchez Crespo, Ph.D.
Universidad de Castilla - La Mancha

Manuel Sánchez Rubio, Ph.D.
Universidad de Alcalá, Universidad Internacional de la Rioja

Lorena Barona López, Ph.D.
Escuela Politécnica Nacional

Leonardo Valdivieso Caraguay, Ph.D.
Escuela Politécnica Nacional

Luis Urquiza Aguiar, Ph.D.
Escuela Politécnica Nacional

CERTIFICATION

I certify that PATRICIO XAVIER ZAMBRANO RODRÍGUEZ has carried out his research under my supervision. To the best of my knowledge, the contributions of this work are novel

PhD. Jenny Gabriela Torres Olmedo

ADVISOR

STATEMENT

I hereby declare under oath that I am the author of this work, which has not previously been presented for obtaining any academic degree or professional qualification. I also declare that I have consulted the bibliographic references included in this document.

Through this declaration, I transfer my intellectual property rights corresponding to this thesis, to the Escuela Politécnica Nacional, as established by the Intellectual Property Law of Ecuador, its Regulations and the current institutional norms.

I declare that this work is based on the following articles of my authorship (as main author or co-author) related to the title of this thesis.

- **P. Zambrano**, J. Torres, L. Tello-Oquendo, A. Yánez, L. Velásquez. "On the modeling of cyber-attacks associated with social engineering: A parental control prototype," *Journal of Information Security and Applications*. 2021. Under review (passed the first round of review). **Journal JCR Q2**

- **P. Zambrano**, J. Torres, Á. Yánez, A. Macas, L. Tello-Oquendo. "Understanding cyberbullying as an information security attack—life cycle modeling," *Annals of Telecommunications*. 2020. https://doi.org/10.1007/s12243-020-00785-0. **Journal JCR Q3**

- **P. Zambrano**, J. Torres; L. Tello-Oquendo; R. Jácome; M. Benalcázar; R. Andrade, "Technical Mapping of the Grooming Anatomy Using Machine Learning Paradigms: An Information Security Approach," *IEEE Access*, vol. 7, pp. 142129-142146, 2019, doi: 10.1109/ACCESS.2019.2942805. **Journal JCR Q1**

- **P. Zambrano**, J. Torres, P. Flores. "How Does Grooming Fit into Social Engineering?," *Bhatia S., Tiwari S., Mishra K., Trivedi M. (eds), International Conference on Computer, Communication and Computational Sciences (IC4S), 2018. Advances in Computer*

Communication and Computational Sciences. Advances in Intelligent Systems and Computing, vol 924. 2019. Springer, Singapore. https://doi.org/10.1007/978-981-13-6861-5_53.

- **P. Zambrano**, M. Sanchez, J. Torres and W. Fuertes, "BotHook: An option against Cyberpedophilia," in *2017 1st Cyber Security in Networking Conference (CSNet), Rio de Janeiro, 2017*, pp. 1-3, doi: 10.1109/CSNET.2017.8241994.

I also declare that I have acknowledged the collaboration of third parties, and the contribution made by other published or unpublished material.

PATRICIO XAVIER ZAMBRANO RODRÍGUEZ

DEDICATION

I dedicate my work to my daughters, Camila and Eva Luciana, for the unconditional love they give me every day.

Patricio Zambrano

ACKNOWLEDGMENTS

To my parents, for their patience and support in my personal and professional development.

To my supervisor, Jenny Torres, for the trust she placed in me during the research process.

To my friends, Carlos Anchundia and Marco Benalcázar, who supported me unconditionally all along my research process.

To my friend, Pamela Flores, for her patience and support.

To Patricio Tello, research collaborator and friend, who was ready to collaborate in everything in the most professional way.

To my students Ángel Yánez and Andrés Jácome who collaborated in the application part of this research.

Thanks.

Patricio Zambrano

TABLE OF CONTENTS

CERTIFICATION II

STATEMENT III

PROLOGUE VII

LIST OF FIGURES XIII

LIST OF TABLES XVI

RESUMEN 1

ABSTRACT 2

1 INTRODUCTION 3

1.1 Problem Statement 5

1.2 Research Motivation 6

1.3 Objectives 6

1.4 Contributions - Articles 7

1.4.1 Chapter 2.- Contextualization of the grooming problem. 7

1.4.2 Chapter 3.- Study of grooming within the context of information security. 9

1.4.3 Chapter 4.- Replication and validation of the process defined for the analysis of grooming, in another case study: cyber bullying. 10

1.4.4 Chapter 5.- Formalization of the process defined for the analysis of grooming and proposal of a prototype of parental control. 12

References 14

2 BOTHOOK: AN OPTION AGAINST CYBERPEDOPHILIA 21

2.1 Introduction 21

2.2 Related Work 23

2.3 BotHook Platform 23

2.3.1 CCAM 24

2.3.2 BOTM 25

2.3.3 PTCM 26

2.4 Analysis and Discussion . . . 27
2.5 Conclusion . . . 28
References . . . 29

3 HOW DOES GROOMING FIT INTO SOCIAL ENGINEERING? 31
3.1 Abstract . . . 31
3.2 Introduction . . . 31
3.3 Background . . . 33
3.3.1 Social Engineering . . . 33
3.3.2 Grooming . . . 34
3.4 Methodology . . . 36
3.5 Results . . . 37
3.5.1 Information gathering . . . 37
3.5.2 Development of relationship . . . 38
3.5.3 Exploitation of relationship . . . 38
3.5.4 Execution to achieve objective . . . 39
3.6 Discussion . . . 39
3.7 Conclusions . . . 40
References . . . 42

4 TECHNICAL MAPPING OF THE GROOMING ANATOMY USING MACHINE LEARNING PARADIGMS: AN INFORMATION SECURITY APPROACH 45
4.1 Abstract . . . 45
4.2 Introduction . . . 46
4.3 What is Grooming? . . . 48
4.3.1 Stages of Grooming . . . 49
4.3.2 Tools and Technologies to detect grooming . . . 51
4.4 Psychological/technical attacker profile . . . 53
4.5 Related Work . . . 58
4.6 Research Methodology . . . 60
4.6.1 Computer attack evaluation . . . 60
4.6.2 Research questions . . . 61
4.6.3 Experimental approach . . . 62
4.7 Experimentation . . . 63
4.7.1 Experiment 1 . . . 63
4.7.2 Experiment 2 . . . 69
4.8 Answering Research Questions . . . 79

4.8.1 With the use of computer learning, can the phases of grooming be determined as a computer attack? . . . 79
4.8.2 Can grooming be considered an attack vector within the APTs? . . . 80
4.8.3 Can the studies related to grooming support future research associated with social engineering? . . . 82
4.9 Conclusions and Future Work . . . 82
References . . . 84

5 UNDERSTANDING CYBERBULLYING AS AN INFORMATION SECURITY ATTACK - LIFE CYCLE MODELING 91
5.1 Abstract . . . 91
5.2 Introduction . . . 92
5.3 Methodology . . . 93
5.4 Problem Understanding . . . 93
5.4.1 Background . . . 94
5.4.2 Scientific contributions to the field of cyberbullying . . . 95
5.4.3 Cyberbullying seen as a Social Engineering attack . . . 97
5.4.4 Computer attack evaluation . . . 98
5.4.5 Research questions . . . 99
5.5 Data Understanding . . . 100
5.5.1 Knowledge bases related to cyberbullying . . . 100
5.6 Data Preparation . . . 101
5.6.1 Software resources. . . . 101
5.6.2 Data construction. . . . 103
5.6.3 Data processing. . . . 105
5.7 Modeling . . . 105
5.7.1 Building the LDA model for two databases . . . 106
5.7.2 Characterizing the topics with linguistic aspects using EMPATH . . . 107
5.7.3 Life cycles of a computer attack applied to cyberbullying . . . 110
5.8 Evaluation . . . 111
5.8.1 Implementing and evaluating the behavior of two machine learning models . . . 113
5.8.2 Answering the research questions . . . 114
5.9 Conclusions and Future Work . . . 115
References . . . 117

6 ON THE MODELING OF CYBER-ATTACKS ASSOCIATED WITH SOCIAL ENGINEERING: A PARENTAL CONTROL PROTOTYPE 127
6.1 Abstract . . . 127
6.2 Introduction . . . 128
6.3 Problem Understanding . . . 129
6.3.1 Background . . . 129
6.3.2 Review of Topic Modeling Applications . . . 132
6.3.3 Research questions . . . 133
6.4 Framework proposal to model cyber-attacks related to Social Engineering . . . 135
6.4.1 Cyber-attack modeling process . . . 135
6.4.2 Application of the modeling process to cyber attacks . . . 137
6.5 Implementation of a parental control prototype . . . 144
6.5.1 Operational component . . . 146
6.5.2 Results' summary . . . 149
6.6 Answering the research questions . . . 150
6.7 Conclusions and Future Work . . . 151
References . . . 153

7 DISCUSSION 162
References . . . 167

8 CONCLUSIONS 169
8.1 Theoretical aspects . . . 170
8.2 Practical Aspects . . . 171
8.3 Methodological Aspects . . . 172
8.4 Future work . . . 173

LIST OF FIGURES

1.1 Thesis Structure 4

2.1 BotHook Platform 24
2.2 CCAM module 25
2.3 BOTM Module 26
2.4 PTCM module 27

3.1 Krombholz et al. Social Engineering Taxonomy 34
3.2 Grooming stages 35
3.3 Summary of Mitnick phases and grooming 40
3.4 Proposed taxonomy and grooming 40

4.1 Graphic scheme that relates a group of numbered researches (more related to research). All researchers use data from Perverted Justice and process the information with a set of technological tools based on their objectives. 52
4.2 Classification of types of attackers and victims [2]. 54
4.3 Psychological/technical profile of a cyber-pedophile. 57
4.4 Life cycles proposed for computer attacks [47, 52, 53]. 61
4.5 Methodology applied to determine the life cycle of grooming. 62
4.6 Word cloud of pre-processed and processed text. 67
4.7 Word cloud by topic - LDA model. 68
4.8 Topics assigned to grooming by LDA. 69
4.9 Analysis of the operation of the LDA model. 70
4.10 Document topic probabilities using the LDA model. 70
4.11 Process to characterize and define the stages of the life cycle of grooming. . 71
4.12 Linguistic categories LIWC. 72
4.13 Assigning codes to data. 73
4.14 Process of data labeling. 73
4.15 Operational concepts (definition) related to the communicational intentions. . 75
4.16 Class distribution histogram. 77
4.17 Model accuracy and test with new data. 78
4.18 CNN network training. 79

4.19 LSTM network training. 80

5.1 Methodology followed in this study based on [2]. 94
5.2 Life cycles proposed for computer attacks [72, 67]. 100
5.3 Scrapy technique. 104
5.4 Rapidminer and R Studio. 104
5.5 Function developed in Matlab for data preprocessing. 105
5.6 Word cloud of pre-processed (raw data) and processed text (cleaned data). Tweets at top and experiences at the bottom. 106
5.7 Topics assigned to cyberbullying by LDA - Data Experiences. 108
5.8 Topics assigned to cyberbullying by LDA - Data Tweets. 109
5.9 Model building with Experiences. 109
5.10 Model building with Tweets. 110
5.11 EMPATH operation. 111
5.12 Operational concepts (definition) related to the linguistic characteristics. . . . 112
5.13 Results of the implementation of two machine learning algorithms. 114

6.1 Cyber-attack modeling process 135
6.2 Subprocess - Attack selection 136
6.3 Subprocess - Data debugging 136
6.4 Subprocess - Topic model application 137
6.5 Subprocess - Model description 138
6.6 Scrapy technique to download short texts (tweets) 139
6.7 Rapidminer and R Studio technique to download experiences of victims . . . 139
6.8 CRISP-DM model applied to cyber attacks 140
6.9 Graphical comparison of the behavior of the coherence values in 4 topic models - Grooming 142
6.10 Graphical comparison of the behavior of the coherence values in 4 topic models - Cyberbullying 143
6.11 Architecture of the parental control prototype 145
6.12 Main components of the prototype 146
6.13 Modeling process: system permissiveness 147
6.14 Modeling process: system permissiveness 148
6.15 Notification 148

7.1 Contributions - Section 1. 164
7.2 Contributions - Section 2. 165
7.3 Contributions - Section 3. 166
7.4 Contributions - Section 4. 167

LIST OF TABLES

1.1 Environment of grooming and cyber-pedophilia 8

4.1 Research Related To The Study Of Grooming 52
4.2 Materials Used in Experimentation 64
4.3 Grooming station 74
4.4 APT Criteria Applied to Grooming 81

5.1 Cyberbullying - Literature review 96
5.2 Scientific evidence that relates cyberbullying to social engineering 99
5.3 Cyberbullying Related Websites 102
5.4 Materials Used in Experimentation 103

6.1 Topic Modeling - Literature review 134
6.2 Application of CRISP-DM to Grooming and Bullying 139
6.3 Topics defined in the analysis tools 140
6.4 Linguistic tools and security models categorization 143
6.5 Summary of previous research 149
6.6 Prototype results 150

8.1 Summary of previous research 169
8.2 Related research precision measurements 172
8.3 Prototype results 172

RESUMEN

En el campo de la seguridad de la información, hay varias áreas de estudio que están en desarrollo. La ingeniería social es una de ellas ya que aborda los retos multidisciplinares de la ciberseguridad. Hoy en día, los ataques asociados a la ingeniería social son diversos y han sido objeto de numerosas investigaciones; sin embargo, los ataques cibernéticos de naturaleza similar al *grooming* han sido excluidos de estos estudios. En la última década se han realizado diversos esfuerzos para comprender la estructura y el enfoque del *grooming* desde el campo de la informática con el uso de algoritmos de aprendizaje computacional. Sin embargo, estos estudios no están alineados con la seguridad de la información y sus propuestas aún son limitadas. Hoy en día, las técnicas psicológicas utilizadas para hostigar, intimidar, amenazar, robar información son más comunes debido al libre acceso a los recursos tecnológicos y la digitalización de las comunicaciones. El problema puntual de investigación radica en la falta de un proceso formal y metodológico que permita estudiar ataques relacionados al acoso cibernético y relacionarlos a conceptos de seguridad de la información. Esta investigación científica presenta cómo se puede analizar un ciberataque relacionado con el acoso en línea y vincularlo con conceptos de la seguridad de la Información a través del método científico y el planteamiento de un proceso procedimental para este fin. El proceso procedimental permitió analizar datos y detectar las intenciones comunicacionales de los atacantes con alta precisión. Adicionalmente, se presenta un prototipo de control parental funcional que soporta el proceso de modelado propuesto.

PALABRAS CLAVE: Grooming, Cyberbullying, Social Engineering, APT.

ABSTRACT

In the field of information security, there are several areas of study that are under development. Social engineering is one of them since it addresses the multidisciplinary challenges of cybersecurity. Today, the attacks associated with social engineering are diverse and have been the subject of numerous research; however, cyberattacks similar to *grooming* have been excluded from these studies. In the last decade, various efforts have been made to understand the structure and approach of *grooming* from the field of computer science with the use of computational learning algorithms. However, these studies are not aligned with information security and their proposals are still limited. Nowadays, the psychological techniques used to harass, intimidate, threaten, steal information are more common due to the free access to technological resources and the digitalization of communications. The specific research problem lies in the lack of a formal and methodological process that allows studying attacks related to online harassment and relating them to information security concepts. This scientific research presents how a cyberattack related to online harassment can be analyzed and linked to information security concepts through a scientific research method and the approach of a procedural process for this purpose. The procedural process made it possible to analyze data and detect the communicational intentions of the attackers with high precision. Additionally, a functional parental control prototype is presented that supports the proposed modeling process.

KEY WORDS: Grooming, Cyberbullying, Social Engineering, APT.

PROLOGUE

The expansion of Internet access has made it possible to shorten distances in terms of social communications. In this human interaction, in the digital environment, many aspects of their behavior and the development of new trends related to social relations have become evident. In this new era, the use of virtual environments has increased, due to the number and variety of applications to which users have access. The purpose of each application is based on its usefulness and can cover aspects such as social interaction (social networks, dating, blogs, chat meetings), entertainment (video games, video streaming), commerce (online shopping, banking), and education (webinars, online classes). However, the use of digital services not only provides positive aspects but also enables the possibility of crime. One type of criminal activity focuses on the violation of people and their emotional stability, and its means of access is usually textual communication. Victims may be targeted on the basis of their age, gender, ethnicity, race, etc. However, the most vulnerable victims are minors who, due to their ignorance and naivety, can be victims of sexual harassment. Inevitably, these types of victims tend to be at the forefront of technological change and are sometimes very active and exposed on social networks.

The thesis focuses its research on grooming, which has been conceptualized as an attack where aggressors create opportunities that allow child sexual abuse. As a first phase, the research begins with a first approach, through an exhaustive literature review, to the social problem and grooming, defining the scope of the study. From this analysis, a study gap was evidenced where grooming is not analyzed from the concepts of information security, particularly social engineering.

Based on this background, we proceeded to theoretically strengthen the procedural relationship between grooming and social engineering. Then, we formalized the knowledge gathered about grooming in a complete study, which describes the different phases of the attack, aspects related to the profile of the attackers, and their communicative intentions. In aspects of validation and replication, the thesis shows the approach of a process that allows analyzing psychological attacks where persuasion or intimidation are used as their mechanisms. Under this perspective, bullying was analyzed in its general context (income data,

life cycle, and communicative intentions). The study consolidate all this knowledge in the development of a prototype of parental control in real-time. Among the main contributions that the study brings to the scientific field we have:

- A psychological/technical profile of attackers in the context of grooming.
- Definition of the grooming life cycle from an information security perspective.
- Consolidation of grooming as an attack vector within the field of social engineering.
- The generalization of a process that allows analyzing different types of attacks, where persuasion or psychological intimidation is evidenced.
- A parental control prototype that analyzes conversations in real-time and categorizes them according to their level of danger in the different phases established in the life cycles.

1 INTRODUCTION

In the context of cybersecurity, social engineering describes a type of attack in which the attacker exploit human vulnerabilities to breach the security goals of cyberspace elements. Recently, attacks on the privacy of children and adolescents through technological means have increased considerably [1, 2, 3, 4, 5, 6]. Some research related to this social problem addressed different topics such as: the study and analysis of the images of children in Peer-to-Peer (P2P) networks, planning of security models in audio-visual devices for child control with access to the Internet, study of vulnerabilities in online video games, development of communication bots for the detection of potential attackers, forensic tools, and analysis of pedophile behavior within instant messages.

Several scientific proposals that have studied the behavior of online attackers stand out in the study of instant messaging, concluding that the most common technique applied by attackers is *grooming*. Grooming has been studied for over a decade and has been conceptualized as a procedural technique used by cybernetic attackers. It is also regarded as an operational concept, whereby an attacker applies affinity search strategies, while acquires information and sexually desensitizes victims to develop relationships that lead to the satisfaction of the needs of the offender or attacker [2, 1, 3, 7, 8].

In order to better understand the use of grooming as an attack technique, we highlight some well-documented and representative studies of grooming [1, 2, 8, 7, 3]. A particularity of these research is that they analyze the texts of real pedophile conversations. These studies are very diverse and range from Neuro Linguistic Programming (NLP), text mining, mathematical models, to a wide range of artificial intelligence techniques in order to determine a profile or pattern of pedophile behavior [9]. It should be noted that there are contributions related to the study of human behavior and information technologies, in isolation without contributing to the aforementioned subject of study [10, 11, 12, 13, 14].

Nevertheless, there is no a formal and methodological process in the literature that studies

grooming in relation to the criteria and concepts of computer security. Thus, the main goal of this research is to standardize a process to analyze and study different social engineering cyberattacks with computer security concepts and see its applicability or transferability to other similar phenomena.

This thesis is organized as follows. Chapters 2 and 3 contextualize the issue of grooming and its relationship to information security concepts by reviewing existing literature on vulnerabilities related to grooming and social engineering. In particular, these chapters allowed the identification of the study gap that led to the development of the thesis objectives. Chapter 4 establishes the source of information to be analyzed, related to grooming, describing the psychological and technical characteristics of the attackers, and their behavioral patterns on the Internet defined in a life cycle, among other aspects. An important aspect of the scientific method is the replication and validation of the defined process. This process, chapter 5, was applied to a phenomenon with similar characteristics to grooming for its assessment and evaluation. According to the results obtained, Chapter 6 formalizes the procedure established in the study of grooming and presents an application that analyzes conversations in real-time and alerts parents if their children are facing an attack associated with the phases of grooming.

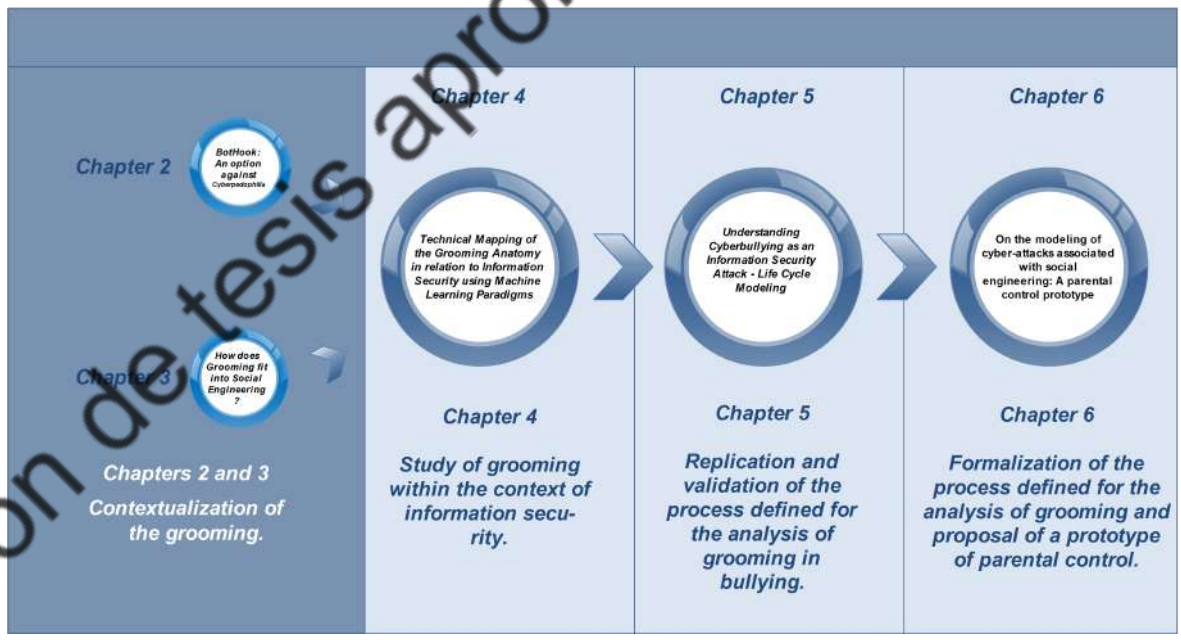


Figure 1.1: Thesis Structure

The thesis contemplated several edges, described in the different chapters, see figure 1.1. These allowed to consolidate the objectives set, based on an exhaustive exploration of grooming, the vulnerability issues surrounding it, and its direct relationship with social enginee-

ring techniques. As a final result, the research provides a procedure that can be replicated in other phenomena of similar characteristics and consolidates its applicability in developments that allow mitigating this social problem in a preventive way.

1.1 PROBLEM STATEMENT

The Internet is certainly an access to opportunities for personal knowledge and interrelation. Nevertheless, nowadays children as well as adolescents are in a variety of ways vulnerable to the dark sides of such technology. A study carried out by UNICEF [15], determined that children and adolescents are connected to Internet, in average, 3 hours per day using social networks. Within this scenario, it remains more than evident that the dignity and reputation of them may be harmed in the cyberspace. In cybersecurity, two types of harassment that affect the dignity of children and adolescents have been distinguished, grooming and cyberbullying. The first harassment is considered as an erotic practice for minors, while the cyberbullying is a process where attackers use technology as a communication medium to threaten, embarrass, intimidate, or criticize their victims.

Studies related to this phenomenon from the fields of computer science and security are still limited [16]. Several factors such as: a) the lack of research based on the direct contact of the victims or attackers, b) the access to specific databases related to this interaction, c) the unification of conceptual criteria that evaluate the nature of the different phenomena associated with grooming, d) the absence of prototypes or proposals that proactively prevent and mitigate these phenomena; among others, could be the cause of the lack of interest of researchers in relating grooming to information security, and creating contributions of scientific relevance [17, 18, 19, 20].

The main research problem lies in the lack of a formal and methodological process that studies grooming in relation to the criteria and concepts of computer security. Based on this, the research raised three research questions:

RQ1. *What is the conceptual and procedural relationship between grooming and social engineering?*

RQ2. *What are the theoretical foundations provided by the study of grooming to the phenomenon associated with cyberbullying?*

RQ3. *How are the different types of online harassment related in their respective contexts or phenomena?*

1.2 RESEARCH MOTIVATION

Pedophilia as a social phenomenon is a psychological disorder where its actors present thoughts, impulses, and sexual fantasies toward minors. The literature refers to medical studies that show neuropsychological characteristics, which are present in different areas of the brain. In relation to their behavior, studies have determined that in most situations they have traits of impulsiveness and difficulty in maintaining interpersonal relationships [21]. It is worth mentioning that these studies are indeterminate and inconclusive to establish the neurobiological profile of the pedophile [22]. To study psychological profiles, a population of pedophiles is required with the predisposition to be studied, an aspect that does not occur in real life. In contrast to online attackers, their behavioral patterns can be analyzed based on the digital information they generate in their attacks.

The motivation of this study is to fill the research gap mentioned before and analyze online attackers through the behavioral patterns that they generate digitally through a formal and methodological process. This process should be standardized in order to analyze and study different social engineering cyberattacks with computer security concepts and see its applicability or transferability to other similar phenomena. Thus justifying its relevance and applications for future research of technical/social interest.

1.3 OBJECTIVES

This thesis aims to *establish the conceptual and procedural relationship between grooming and social engineering.* In order to achieve this goal, some specific objectives are:

- ❖ Consider the use of computer algorithms for the discovery of linguistic characteristics and communicational intentions intrinsic to the phenomenon of grooming applying scientific rigor through the scientific method.
- ❖ Identify the psychological profile of the cyber pedophile and the behavioral patterns they develop in their attacks and analyze them within the context of information security

theory.

- ❖ Explore the impact of the grooming attack on its victims and establish potential strategies to mitigate this social problem.

1.4 CONTRIBUTIONS - ARTICLES

In this section we summarize the contributions of this thesis specifying how the problem is addressed in each article, and the publication status.

1.4.1 Chapter 2.- Contextualization of the grooming problem.

Considering the recommendations described by Barbara Kitchenham [[23] in the field of systematic reviews, a methodological process was followed based on the analysis of key words linked to the phenomenon studied. In this way, search chains were strategically constructed to create criteria for the selection and exclusion of research in search of a study gap. In this exhaustive analysis of the scientific contributions related to the phenomenon conceptualized as grooming, see Table 1.1, five fields of knowledge were established in the scientific-technical field. The first field points to contributions focused on the analysis of videos, photos and other aspects that may threaten the integrity of minors. In the second field, experts in the field propose different psychological profiles of aggressors depending on their approach to the study. The third field includes studies that analyze the different media associated with the Internet, such as social networks and video games, among others. The fourth field describes various scientific contributions aimed at mitigating the risk of attack. The last field includes contributions related to the evaluation of grooming from the point of view of written communications through chats in theoretical-practical aspects with proposals that employ artificial intelligence techniques to determine the communicational aspects of the attackers. It should be noted that the root of the thesis is developed in field 5, however, it is linked to fields 2, 3, and 4, thus achieving transversality in knowledge.

The research in this thesis is focused on the study of grooming, starting with a first approach to the social problem of online attacks, defining the scope of the study, to continue with the theoretical underpinning of grooming and social engineering.

The first work entitled "BotHook: An option against Cyberpedophilia"[9] under the modality

Table 1.1: Environment of grooming and cyber-pedophilia

Field of knowledge 1	Field of knowledge 2	Field of knowledge 3	Field of knowledge 4	Field of knowledge 5
Research related to child abuse	Proposals for aggressor profiles	Analysis of Internet communications media	Process proposals to mitigate risks	Research related to text analysis with intelligent systems
[24, 25, 26, 27, 28, 29] [30, 31, 32, 33, 34, 35] [36, 37, 38]	[24, 39, 40, 1, 41, 42] [43, 44, 45]	[39, 46, 47, 7, 42, 48]	[49, 33, 17, 50, 51, 52] [43, 45, 53, 54]	[5, 1, 2, 39, 4, 55, 8, 56] [57, 20, 58, 51, 52, 43] [48, 59, 44, 60, 61, 53, 62]

Keywords: Child online safety, Internet safety, child sexual abuse, child pornography, nudity, Paedophile, pedophile, sexual predation, Grooming, Sexual offender, Online Child Grooming, Predatory

of Poster, included an exhaustive systematic review of the literature that revolved around online pedophilia, grooming, and other keywords that relate research to this social phenomenon developed in cyberspace. Under this context, proposals in different fields of study were evidenced, such as psychological-social, technological, and educational, among others. From the point of view of technological contributions, many of them based their proposals on Artificial Intelligence (AI) for the analysis of different types of information. In the article, a technological proposal "bot" was developed for interaction with potential pedophiles to obtain their communication patterns. This proposed platform use AI techniques, natural language processing, and the classification of psychological information for its development. This mechanism was proposed in order to reduce cyber-attacks against the innocence, integrity, and security of minors, who are exposed to the Internet. In the development of this research it became evident that the different contributions were not aligned to concepts and criteria proper to information security, this gap was developed in the publication entitled: "How Does Grooming Fit into Social Engineering?"[16].

Conceptual aspects of information security such as assets, attack vectors, and mechanisms, as well as principles such as confidentiality, integrity, and availability were developed in this research. In the particular case of grooming, people and their emotional stability were established as assets, as well as the most violated principles such as confidentiality and integrity. Being a social phenomenon concerning human behavior, social engineering was the most related context and we proceeded to study the various contributions to this field. Once the research was completed, the results allowed us to include grooming within the processes of social engineering, validating its insertion with the phases of social engineering designed by Kevin Mitnick. Scientific evidence showed that grooming is the technique applied by sexual harassers of minors. However, researchers do not interrelate the contributions generated by grooming and social engineering, which allows inferring that the studies are isolated and incomplete and should be consolidated with complementary scientific evidence.

The adaptation of grooming to the context of social engineering can be used in the scientific

area in a variety of ways. From an academic point of view, this research presented a starting point for security researchers trying to generate contributions to social problems that develop on the Internet. From a computational point of view, applications can be generated that facilitate the preventive detection of possible attacks on children and adolescents through computer learning.

The contributions were published in:

- **Conference:** *1st Cyber Security in Networking Conference (CSNet), 2017.*
- **Conference:** *International Conference on Computer, Communication and Computational Sciences (IC4S), 2018.*

1.4.2 Chapter 3.- Study of grooming within the context of information security.

With the background obtained in the research, we proceeded to formalize the knowledge obtained about grooming in a complete study, which describes the different phases of the attack, aspects related to the profile of the attackers, and their communicational intentions. In relation to the contributions aligned to the technological field, several efforts were made to understand the structure and approach of grooming. However, these studies are not aligned with information security. Part of these contributions developed their proposals using computational learning algorithms, with the purpose of recognizing linguistic characteristics and patterns in order to recognize attackers'behavioral characteristics.

One of the objectives of the research is to analyze and characterize the communicational intentions of the attackers. In order to achieve this objective, it was required the selection of a data mining methodology, a database related to the field of study, a mechanism of classification of these in groups of words to identify the different phases through which the victims are attacked, to finally contrast these phases with concepts of information security. The steps described in the CRISP-DM methodology were followed to obtain, process, and clean the database. The database considered and collected for the research was obtained from Perverted-Justice (PJ), a website that stores pedophile and pseudo-victim chats. Regarding the data classification mechanism, a generative LDA model was selected to understand and explain the communicational intentions inherent in the attackers'chat texts in each grouping.

Finally, these clusters were categorized with information security concepts (life cycle phases of advanced persistent threats APTs). As a relevant aspect of the study, supervised and unsupervised learning algorithms were implemented to evaluate future categorizations of texts not described in the databases, with the aim of developing a prototype of real-time parental control in the future.

The main contributions of this study are summarized as follows:

- A psychological and technical profile of the type of attacker associated with online pedophilia is presented;
- Grooming as a vector of attack within social engineering and computer security is positioned; this will allow supporting research related to determining patterns of malicious behavior online;
- Through topic modeling, different stages of the life cycle of an attack associated with social engineering are determined; The analysis of the data, under this consideration, allows the aspects of subjectivity, in the selection of the number of phases of the life cycles, are not a determining factor in the results of their studies.
- Application of a linear machine learning algorithm to classify binding texts to the study area. The research addresses an application context, under the training of an AI model, which allows classifying new instant messaging texts in the stages of the defined life cycle, thus proposing a proactive process in the early detection of future attacks.

This contribution was published in the **Journal:** *IEEE Access, 2019. JCR Q1.*

1.4.3 Chapter 4.- Replication and validation of the process defined for the analysis of grooming, in another case study: cyber bullying.

As part of a validation and replication process, a new phenomenon with similar characteristics to grooming was analyzed, *cyber bullying*. In this analysis, the input data (short texts and experiences) and the procedure described in grooming were evaluated.

Like grooming, cyber bullying is a phenomenon of social interaction that is reflected in physical and virtual spaces (Internet - social networks, chats, etc.). Cyber bullying has a particularity in relation to grooming, in this attack, the victims and attackers are of similar age, while in grooming the bully is an adult. When dealing with the purpose of the attacks, they are completely different, while the cyber bullying attacker seeks to humiliate the victims, the grooming attacker tries to abuse them sexually physically. For this reason, the research is not based on the purpose of the attack but on the study of the attack process, taking as a starting point studies related to this phenomenon.

In the first analysis, the study determined that cases of cyber bullying are becoming more frequent due to free access to technological resources and the new trends of social interrelation in cyberspace. The most characteristic trends are social networks, online games, and virtual classes. However, studies related to this phenomenon are still limited, since, like grooming, they are not analyzed from the perspective of information security.

In terms of replication and validation, this research applies the grooming analysis process and proposes a life cycle model of cyber bullying through the modeling of topics conceptualizing the different stages of this. For the selection and treatment of the data (short texts), this study relied on a specific literature review, where the technical procedures that allowed obtaining the knowledge bases from the experiences of victims of cyber bullying and tweets of attackers were collected.

Studying this phenomenon from the point of view of information security, such as social engineering, made it possible to link the techniques, methodologies, and other procedural aspects proposed in grooming (replication). With this linkage, more specialized knowledge on the subject is consolidated, thus avoiding incomplete and dispersed proposals. The results obtained in this research justify the validity of the established process, allowing its applicability in the study of other phenomena associated with cyber bullying, such as mobbing, and gender violence, among others.

The main contributions of this study are summarized as follows:

- ❖ Research was replicated in aspects related to obtaining data from short texts, thus validating the use of this type of groomimg data.

- ❖ Correlation of the grooming analysis process applied to cyber bullying with longer texts related to the victimséxperiences.

- ❖ The process of the analysis done for grooming is replicated and validated with a similar phenomenon: cyber bullying.

- ❖ Correlation of the grooming analysis process applied to cyber bullying with a specialized linguistic software (EMPATH), to determine the communicational intentions of the attackers.

This contribution was published in the **Journal:** *Annals of Telecommunications, 2020, JCR Q3.*

1.4.4 Chapter 5.- Formalization of the process defined for the analysis of grooming and proposal of a prototype of parental control.

Finally, this research formalizes the process of modeling cyber-attacks related to any type of cyber bullying, defining its stages through topic modeling. This process was applied to model cyber-attacks associated with grooming and cyber bullying, in which the use of psychological manipulation techniques by the attackers was evidenced. Another important aspect of the research, in relation to innovation, was the implementation of a parental control prototype, which in real time can detect and alert of suspicious conversations related to grooming and cyber bullying. The basis of this prototype was developed from the machine learning technique used in theoretical/practical research on grooming and cyber bullying.

The main contributions of this study are summarized as follows:

- ❖ A framework for modeling cyber-attacks related to cyber bullying and grooming. This proposal aims to standardize and link future studies in aspects related to online human behavior.

- ❖ Present online data collection techniques and treatment of these in the context of instant messaging. This contribution allows the reader to analyze current techniques to collect, cleanse, and process unstructured data that resides on the Internet.

- ❖ Present alternative models of topics in short text processing, in order to determine groupings of words that describe the linguistic context of the phenomenon studied.

- ❖ Present functional tools that allow establishing lexical meanings to the topics obtained in the modeling process.
- ❖ Show the architecture and operation of a parental control prototype based on the proposed framework.

This contribution passed the first round of review in the **Journal:** *Journal of Information Security and Applications, 2021. JCR Q2.*

REFERENCES

[1] Lyta Penna, Andrew Clark, and George Mohay. Challenges of automating the detection of paedophile activity on the internet. *Proceedings - First International Workshop on Systematic Approaches to Digital Forensic Engineering*, 2005:206–220, 2005.

[2] Nick Pendar. Toward spotting the pedophile telling victim from predator in text chats. *ICSC 2007 International Conference on Semantic Computing*, (c):235–241, 2007.

[3] Ryan C.W. Hall and Richard C.W. Hall. A profile of pedophilia: Definition, characteristics of offenders, recidivism, treatment outcomes, and forensic issues. *Mayo Clinic Proceedings*, 82(4):457–471, 2007.

[4] Dasha Bogdanova, Saint Petersburg, Paolo Rosso, and Thamar Solorio. On the impact of sentiment and emotion based features in detecting online sexual predators. *Proceedings of the 3rd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, (July):110–118, 2012.

[5] Keith F Durkin. Misuse of the internet by pedophiles: Implications for law enforcement and probation practice. *Fed. Probation*, 61:14, 1997.

[6] Hugo Jair Escalante, Esaú Villatoro-Tello, Sara E Garza, A Pastor López-Monroy, Manuel Montes-y Gómez, and Luis Villaseñor-Pineda. Early detection of deception and aggressiveness using profile-based representations. *Expert Systems with Applications*, 89:99–111, 2017.

[7] India McGhee, Jennifer Bayzick, April Kontostathis, Lynne Edwards, Alexandra McBride, and Emma Jakubowski. Learning to identify internet sexual predation. *International Journal of Electronic Commerce*, 15(3):103–122, 2011.

[8] A.E. Cano, M. Fernandez, and H. Alani. Detecting child grooming behaviour patterns on social media. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8851:412–427, 2014.

[9] Patricio Zambrano, Marco Sanchez, Jenny Torres, and Walter Fuertes. Bothook: An option against cyberpedophilia. In *Cyber Security in Networking Conference (CSNet), 2017 1st*, pages 1–3. IEEE, 2017.

[10] Katharina Krombholz, Heidelinde Hobel, Markus Huber, and Edgar Weippl. Advanced social engineering attacks. *Journal of Information Security and Applications*, 22:113–122, 2015.

[11] Joseph M. Hatfield. Social engineering in cybersecurity: The evolution of a concept. *Computers and Security*, 73:102–113, 2018.

[12] Taimur Bakhshi. Social engineering: revisiting end-user awareness and susceptibility to classic attack vectors. In *Emerging Technologies (ICET), 2017 13th International Conference on*, pages 1–6. IEEE, 2017.

[13] Samar Albladi and George RS Weir. Vulnerability to social engineering in social networks: a proposed user-centric framework. In *Cybercrime and Computer Forensic (ICCCF), IEEE International Conference on*, pages 1–6. IEEE, 2016.

[14] Surbhi Gupta, Abhishek Singhal, and Akanksha Kapoor. A literature survey on social engineering attacks: Phishing attack. In *Computing, Communication and Automation (ICCCA), 2016 International Conference on*, pages 537–540. IEEE, 2016.

[15] Dante Contreras and Paulina Sepúlveda. Effect of lengthening the school day on mother's labor supply. *The World Bank Economic Review*, 2016.

[16] Patricio Zambrano, Jenny Torres, and Pamela Flores. How does grooming fit into social engineering? In *Advances in Computer Communication and Computational Sciences*, pages 629–639. Springer, 2019.

[17] R O'Connel. Typology of cybersex exploitation and online grooming process. *Cyberspace Research Unit, University of Central Lancashire, the United Kingdom, Tech. Rep.*, 2014.

[18] Dimitrios Michalopoulos and Ioannis Mavridis. Utilizing document classification for grooming attack recognition. *Proceedings - IEEE Symposium on Computers and Communications*, pages 864–869, 2011.

[19] Nick Pendar. Toward spotting the pedophile telling victim from predator in text chats. In *International Conference on Semantic Computing (ICSC 2007)*, pages 235–241. IEEE, 2007.

[20] Fergyanto E Gunawan, Livia Ashianti, Sevenpri Candra, and Benfano Soewito. Detecting online child grooming conversation. In *2016 11th International Conference on Knowledge, Information and Creativity Support Systems (KICSS)*, pages 1–6. IEEE, 2016.

[21] Agustin Malón. Pedophilia: A diagnosis in search of a disorder. *Archives of sexual behavior*, 41(5):1083–1097, 2012.

[22] Giulio Perrotta. Pedophilia: definition, classifications, criminological and neurobiological profiles, and clinical treatments. a complete review. *Open Journal of Pediatrics and Child Health*, 5(1):019–026, 2020.

[23] Barbara Kitchenham. Procedures for performing systematic reviews. *Keele, UK, Keele University*, 33(2004):1–26, 2004.

[24] Siew Yong, Dale Lindskog, Ron Ruhl, and Pavol Zavarsky. Risk mitigation strategies for mobile Wi-Fi robot toys from online pedophiles. *Proceedings - 2011 IEEE International Conference on Privacy, Security, Risk and Trust and IEEE International Conference on Social Computing, PASSAT/SocialCom 2011*, pages 1220–1223, 2011.

[25] Mateus De Castro Polastro and Pedro Monteiro Da Silva Eleuterio. A statistical approach for identifying videos of child pornography at crime scenes. *Proceedings - 2012 7th International Conference on Availability, Reliability and Security, ARES 2012*, pages 604–612, 2012.

[26] Moshe Rutgaizer, Yuval Shavitt, Omer Vertman, and Noa Zilberman. Detecting pedophile activity in BitTorrent networks. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7192 LNCS:106–115, 2012.

[27] Matthieu Latapy, Clémence Magnien, and Raphaël Fournier. Quantifying paedophile activity in a large P2P system. *Information Processing and Management*, 49(1):248–263, 2013.

[28] Kathryn C Seigfried-Spellar. Measuring the Preference of Image Content for Self-reported Consumers of Child Pornography. *Digital Forensics and Cyber Crime, Icdf2C 2012*, 114:81–90, 2013.

[29] Alfred Hofmann, Ursula Barth, Ingrid Haas, and Frank Holzwarth. Detection of Child Sexual Abuse Media : Classification of the Associated Filenames. pages 1–5, 2013.

[30] Riccardo Satta, Javier Galbally, and Laurent Beslay. Children gender recognition under unconstrained conditions based on contextual information. *Proceedings - International Conference on Pattern Recognition*, pages 357–362, 2014.

[31] Natalie Bennett and William T O Donohue. Forensic Interviews Regarding Child Sexual Abuse. pages 307–316, 2016.

[32] Vassilios Chatzis, Fotios Panagiotopoulos, and Vassilios Mardiris. Face to Iris Area Ratio as a feature for children detection in digital forensics applications. *2016 Digital Media Industry and Academic Forum, DMIAF 2016 - Proceedings*, pages 121–124, 2016.

[33] Noor Maizura Mohamad Noor and Salwana Mohamad @ Asmara. Intelligent interpretation and analysis of child sexual abuse forensic evidence: A preliminary study. *2010 International Symposium on Information Technology*, pages 624–629, 2010.

[34] Ding Peng, Dai Guilan, and Zhang Yong. Contextual-lda: a context coherent latent topic model for mining large corpora. In *2016 IEEE Second International Conference on Multimedia Big Data (BigMM)*, pages 420–425. IEEE, 2016.

[35] Katrinna MacFarlane and Violeta Holmes. Agent-mediated information exchange: Child safety online. *Proceedings - International Conference on Management and Service Science, MASS 2009*, 2009.

[36] Janis Wolak, David Finkelhor, Kimberly J Mitchell, and Michele L Ybarra. Online “predators” and their victims: Myths, realities, and implications for prevention and treatment. 2010.

[37] Neelam Saleem Khan, Mohammad Ahsan Chishti, and Mahreen Saleem. Identifying various risks in cyber-security and providing a mind-map of network security issues to mitigate cyber-crimes. In *Proceedings of 2nd International Conference on Communication, Computing and Networking*, pages 93–103. Springer, 2019.

[38] Mario Ron, Oswaldo Rivera, Walter Fuertes, Theofilos Toulkeridis, and Javier Díaz. Cybersecurity baseline, an exploration, which permits to delineate national cybersecurity strategy in ecuador. In *International Conference on Information Technology & Systems*, pages 847–857. Springer, 2019.

[39] Paul Elzinga, Karl Erich Wolff, and Jonas Poelmans. Analyzing chat conversations of pedophiles with temporal relational semantic systems. *Proceedings - 2012 European Intelligence and Security Informatics Conference, EISIC 2012*, pages 242–249, 2012.

[40] Dasha Bogdanova, Paolo Rosso, and Thamar Solorio. On the impact of sentiment and emotion based features in detecting online sexual predators. In *Proceedings of the 3rd workshop in computational approaches to subjectivity and sentiment analysis*, pages 110–118. Association for Computational Linguistics, 2012.

[41] Munish Chopra, Miguel Vargas Martin, Luis Rueda, and Patrick CK Hung. Toward new paradigms to combating internet child pornography. In *2006 Canadian Conference on Electrical and Computer Engineering*, pages 1012–1015. IEEE, 2006.

[42] Ritika Pandey and George O Mohler. Evaluation of crime topic models: topic coherence vs spatial crime concentration. In *2018 IEEE International Conference on Intelligence and Security Informatics (ISI)*, pages 76–78. IEEE, 2018.

[43] C.H Ngejane, G Mabuza-Hocquet, J.H.P Eloff, and S Lefophane. Mitigating online sexual grooming cybercrime on social media using machine learning: A desktop survey. In *2018 International Conference on Advances in Big Data, Computing and Data Communication Systems (icABCD)*, pages 1–6, 2018.

[44] Muhammad Ali Fauzi and Patrick Bours. Ensemble method for sexual predators identification in online chats. In *2020 8th International Workshop on Biometrics and Forensics (IWBF)*, pages 1–6, 2020.

[45] Sara Ashry and Walid Gomaa. Descriptors for human activity recognition. In *2019 7th International Japan-Africa Conference on Electronics, Communications, and Computations, (JAC-ECC)*, pages 116–119, 2019.

[46] Dasha Bogdanova, Paolo Rosso, and Thamar Solorio. Exploring high-level features for detecting cyberpedophilia. *Computer Speech and Language*, 28(1):108–120, 2014.

[47] Peter Bourgonje, Julian Moreno-Schneider, Ankit Srivastava, and Georg Rehm. Automatic classification of abusive language and personal attacks in various forms of online

communication. In *International Conference of the German Society for Computational Linguistics and Language Technology*, pages 180–191. Springer, Cham, 2017.

[48] Zheming Zuo, Jie Li, Philip Anderson, Longzhi Yang, and Nitin Naik. Grooming detection using fuzzy-rough feature selection and text classification. In *2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–8, 2018.

[49] Lyta Penna, Andrew Clark, and George Mohay. A framework for improved adolescent and child safety in MMOs. *Proceedings - 2010 International Conference on Advances in Social Network Analysis and Mining, ASONAM 2010*, pages 33–40, 2010.

[50] Surbhi Gupta, Abhishek Singhal, and Akanksha Kapoor. A literature survey on social engineering attacks: Phishing attack. In *2016 international conference on computing, communication and automation (ICCCA)*, pages 537–540. IEEE, 2016.

[51] Aishwarya Upadhyay, Akshay Chaudhari, Arunesh, Sarita Ghale, and S. S. Pawar. Detection and prevention measures for cyberbullying and online grooming. In *2017 International Conference on Inventive Systems and Control (ICISC)*, pages 1–4, 2017.

[52] Philip Anderson, Zheming Zuo, Longzhi Yang, and Yanpeng Qu. An intelligent online grooming detection system using ai technologies. In *2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–6, 2019.

[53] Marsela Nur Rita and Fungai Bhunu Shava. Chatbot driven web-based platform for online safety and sexual exploitation awareness and reporting in namibia. In *2021 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)*, pages 1–5, 2021.

[54] Aria Ghora Prabono, Seok-Lyong Lee, and Bernardo Nugroho Yahya. Context-based similarity measure on human behavior pattern analysis. *Soft Computing*, 23(14):5455–5467, 2019.

[55] Anna Vartapetiance and Lee Gillam. “Our Little Secret”: pinpointing potential predators. *Security Informatics*, 3(1):1–19, 2014.

[56] Dasha Bogdanova, Paolo Rosso, and Thamar Solorio. Modelling Fixated Discourse in Chats with Cyberpedophiles. *Proceedings of the Workshop on Computational Approaches to Deception Detection*, pages 86–90, 2012.

[57] Hady Pranoto, Fergyanto E. Gunawan, and Benfano Soewito. Logistic Models for Classifying Online Grooming Conversation. *Procedia Computer Science*, 59(Iccsci):357–365, 2015.

[58] Seyed Ali Bahrainian, Ida Mele, and Fabio Crestani. Modeling discrete dynamic topics. In *Proceedings of the Symposium on Applied Computing*, pages 858–865, 2017.

[59] Patrick Bours and Halvor Kulsrud. Detection of cyber grooming in online conversation. In *2019 IEEE International Workshop on Information Forensics and Security (WIFS)*, pages 1–6, 2019.

[60] Parisa Rezaee Borj and Patrick Bours. Predatory conversation detection. In *2019 International Conference on Cyber Security for Emerging Technologies (CSET)*, pages 1–6, 2019.

[61] H.M. Rafi Hasan, AKM Shahariar Azad Rabby, Mohammad Touhidul Islam, and Syed Akhter Hossain. Machine learning algorithm for student's performance prediction. In *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pages 1–7, 2019.

[62] Yash Singla. Detecting sexually predatory behavior on open-access online forums. In *Proceedings of Research and Applications in Artificial Intelligence*, pages 27–40. Springer, 2021.

2 BOTHOOK: AN OPTION AGAINST CYBERPEDOPHILIA

Patricio Zambrano[1], Marco Sánchez[1], Jenny Torres[1], Walter Fuertes[1,2]

[1]Faculty of System Engineering, Escuela Politécnica Nacional, Quito, Ecuador

[2]Universidad de las Fuerzas Armadas ESPE, Sangolquí,

Email: {patricio.zambrano, marco.sanchez01, jenny.torres}@epn.edu.ec,
wmfuertes@espe.edu.ec

ABSTRACT

This study presents BotHook, a cyber pedophile trend characterization platform and a hook for computer criminals on the Internet. This work in progress represents a proposal of a distributed platform (chatbot) that includes a module of attraction of pedophile interest, an intelligent engine of a question-answer analysis, and an automatic characterization of pedophile trends. In order to perform this research, we carried out a systematic review of the related work in the existing literature, with the purpose of establishing our own framework. In this context, we propose to use artificial intelligence, natural language processing and classification of psychological information to improve the current chatbot platforms. This proposal encourages to contribute with the field of cybersecurity, in the reduction of cyberattacks against the innocence, integrity and safety of adolescents, who may be exposed on the Internet.

KEY WORDS: Cyberpedophile, chatbot, cybersecurity.

2.1 INTRODUCTION

The Internet is certainly an access to opportunities for personal knowledge and interrelation. Nevertheless, nowadays children as well as adolescents are in a variety of ways vulnerable

to the dark sides of such technology. A study carried out by UNICEF [1], determined that children and adolescents are connected to the Internet 3 hours per day, in average, using social networks. Within this scenario, it remains more than evident that the dignity and reputation of them may be harmed in the cyberspace. Viola [2] defines classic traits of pedophiles, such as gender, age, marital status, affinity for sharing spaces with infants, introverted personality, continuous search for jobs related with infants, work as volunteers, among others. The main difference between a pedophile and a cyber pedophile is the means of access to infants through open media such as the Internet. In cybersecurity two types of harassment that affect children's and adolescent's dignity have been distinguished, grooming and cyber pedophilia. The first harassment is considered as an erotic practice for minors, while the cyber pedophilia is a behavior that consists of sexual intimidation through the media [3].

In 2016, research conducted by Europol [4], evidenced a network of Internet pedophiles with up to 70,000 members who were part of an Internet forum. That has been probably the world's largest online pedophile network. In the same year, the dutch organization for children's and adolescent's fundamental rights, Terre des hommes, with the project Sweetie (a 9-year-old virtual girl) successfully detected one thousand individuals who implicitly asked Sweetie for sexual acts in exchange for money [5]. Between 2014 and 2016 two platforms of chatbot [6, 7] were proposed. These platforms collaborated in the early detection of cyber pedophiles, nevertheless these solutions need to be discussed and improved in aspects of classification, gender language and even slang, considering the region where the platform may be applied.

Based on the described scenario, in this study we present BotHook, a conceptual platform which focuses on studying and improving the aforementioned platforms by applying a better selection of the vulnerable group, in order to set up standardized knowledge bases with the help of specialized psychologists. Additionally, the platform includes a module with Natural Language Processing (NLP) and Artificial Intelligence (AI) techniques, which allows to obtain more assertive responses by the system and in parallel to classify into categories alleged aggressors based on the methods proposed by specialists. Finally, BotHook will also be able to detect cybercriminals who try to exploit vulnerabilities of the system. The paper proceeds as follows. Section II overviews the related studies. Section III details the structure of BotHook Platform. Section IV includes an analysis and discussion of the functionality of the platform. Section V concludes the paper.

2.2 RELATED WORK

In the literature, regarding cybersecurity, there are few contributions in the early detection of cyberpedophiles through chatbot [6], [7]. Callejas-Rodríıguez et al. [6] describe an approach for the programming of a chatbot, which has as its main characteristic an adolescent personality, able to dialogue in spanish language (Mexican). The main goal of this proposal is to contribute in the field of prevention of cyberpedophilia. Carlos Laorden et al. [7] present the Negobot project, which is a contribution against cyberpedophilia and which purpose is to represent a child in chats, social networks and other channels susceptible to pedophile attacks. Negobot is developed with NLP, information retrieval and AI. Nevertheless, the main contribution is the application of game theory. In this context, Negobot proposes, first of all, a competitive game where the system identifies the best strategies to achieve its objective, obtaining information that allows it to infer if the subject involved in the conversation with the chatbot has pedophile tendencies or not. The authors say that Negobot is the first chatbot system based on game theory to detect pedophile tendencies, analyzing the conversations in real time and evaluating the subjects that interact with it.

As explained before, the research area invests a lot in the improvement of these applications (chatbots) that point to specific functionalities, enabled by tools such as ALICE (Artificial Linguistic Internet Computer Entity). It is a free distribution of chatbot based on AIML, open language, minimalist and with stimulus-response used to create bot personalities [8]. ALICE will be considered as the basis for a next phase, the experimental phase, of the platform proposed in this paper.

2.3 BOTHOOK PLATFORM

The BotHook platform includes 3 principal modules: 1) CCAM capture, classification and analysis of cybercriminals, both, cyber pedophiles and criminals which intend to infringe on the system to trade with the retrieved information, 2) BOTM a chatbot module which includes aspects of specificity and; 3) PTCM a pedophile trend characterization module. Figure 1 represents the BotHook platform.

2.3.1 CCAM

It is developed to attract and analyze cyber attacks that intend to violate the system. These attacks are classified into classics (exploration of system failures) and social engineering that make use of the experimental platform. As shown in Figure 2, it has 3 sub modules: DAS, CPHOOK and PPTC.

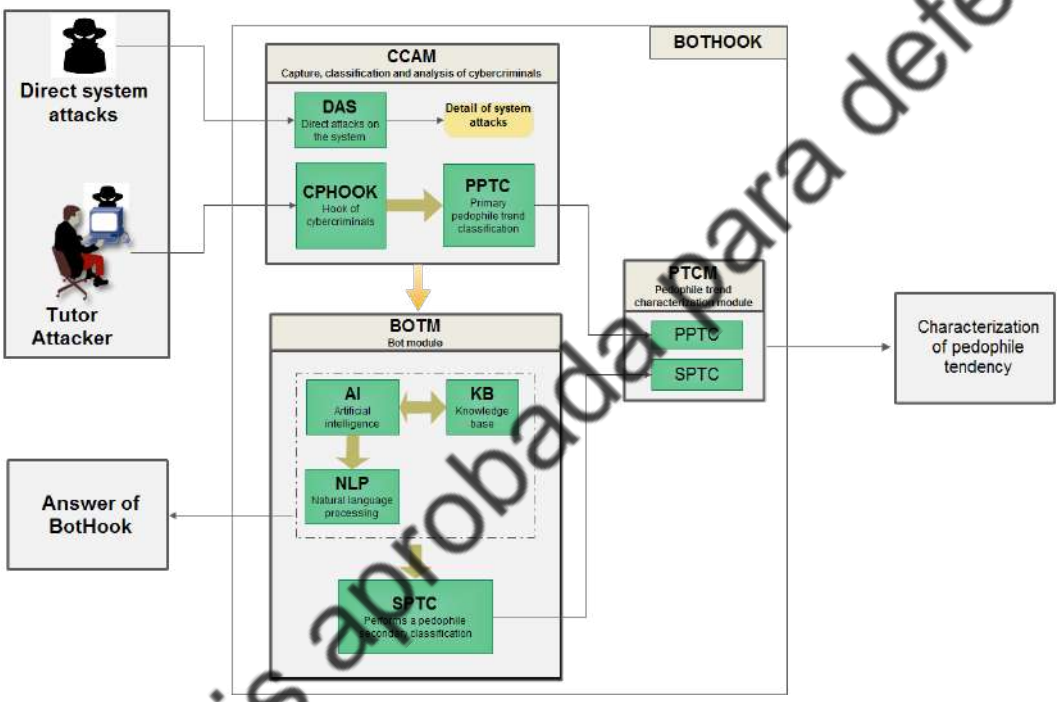


Figure 2.1: BotHook Platform

DAS is a sub module that collects information from direct attacks on the system. This sub module is characterized by collecting all the information necessary to classify potential computer criminals, thus determining their tools, attack tactics and relevant information that characterizes them.

CPHOOK is a sub module that shows a functional application that attracts or serves as a hook for cybercriminals. This sub module is structured by a web portal with a non-profit domain, where students of 12 years old want to learn or reinforce their knowledge in the English language.

PPTC is a sub module where the primary pedophile trend classification is done.

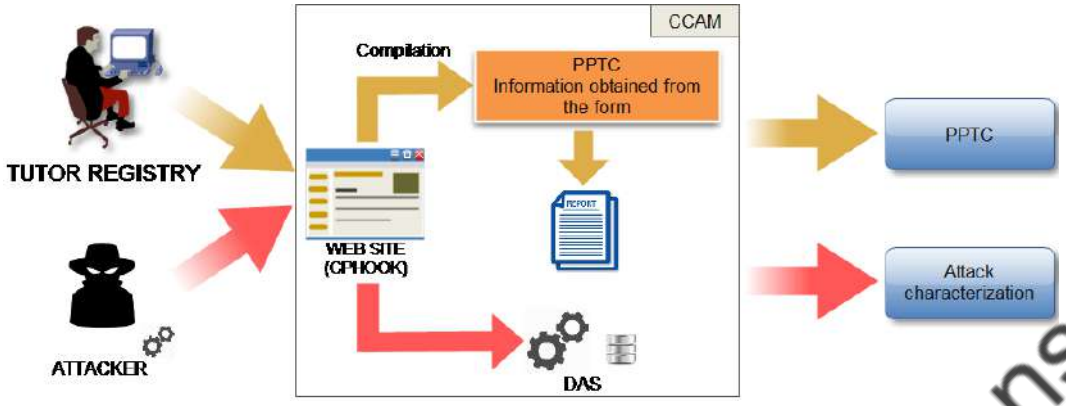


Figure 2.2: CCAM module

2.3.2 BOTM

In communicational aspects, the dialogue established between the system and the potential attacker must be interactive in the construction of the discussed topic without any possibility of expose some system error that alerts the potential attacker and forces him to leave the conversation. As shown in Figure 3, this module develops four main sub modules:

AI is a sub module that executes the function of training by predicting the next dialogue in a given conversational context. As an example, the SEQ2SEQ model of AI in combination with the maximum-likelihood estimation (MLE) achieves this goal [9]. The use of this model in combination with others, offers greater certainty in the response that the sub module NLP shows to the attacker.

NLP, is a sub module that uses tools with the Natural Language ToolKit (NLTK). NLTK is used to split words in a string of text and separate the text into parts of speech by tagging word labels according to their positions and functions in the sentence. The resulting tagged words are then processed to extract the meaning and produce a response. Different grammar rules are used to categorise the tagged words in the text into groups or phrases relating to their positions.

KB, is the knowledge base sub module, with the conversations obtained in the research. They are based on previously established cross-questionnaire (question-answer).

SPTC is a sub module that works in combination with AI, which is in charge of recognizing patterns of potential attackers, assigning categories of danger. These categories are developed in conjunction with specialist psychologists.

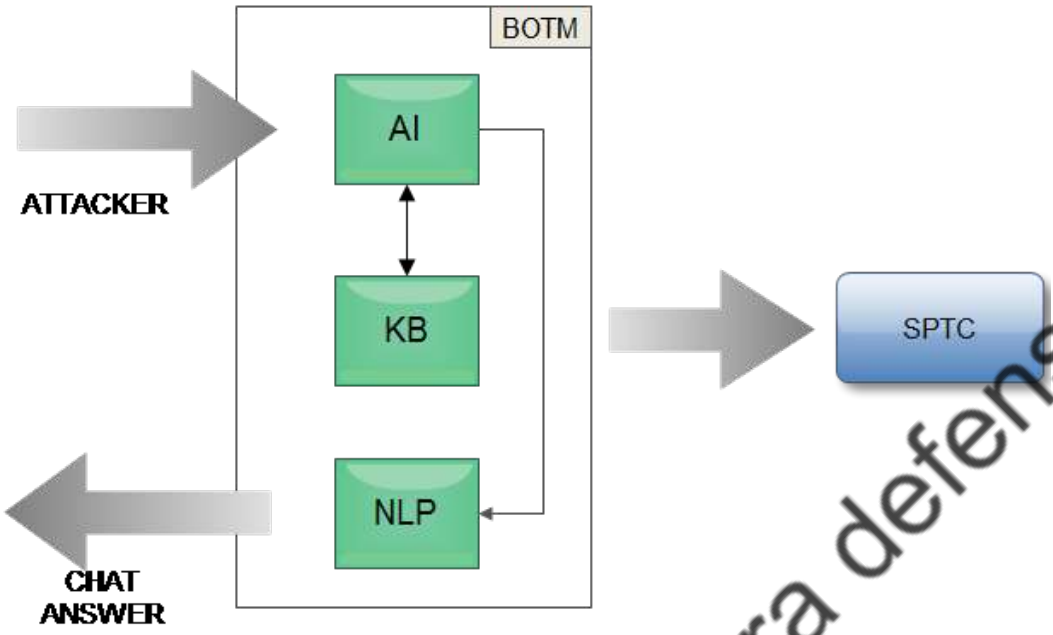


Figure 2.3: BOTM Module

BOTM is projected to have the ability to analyze, classify and respond appropriately to any established conversation. This module differs from the work found in literature, because of its specificity of age, gender, socio-economic status and student slang. Considering the research in [10], the system needs to be inquisitive. However, in case of not being able to give a response, it musk ask new questions until obtaining a coherent one that could be interpreted by the system. This answer will go through a classification processes with AI for system's learning in response to future questions and NLP for the all the system's replies. The knowledge base should be elaborated from a process of collection of questions-answers in a group of 12 years old male and female students, in public and private establishments. With this, it will be possible to standardize adolescent slang and differentiate responses by gender.

2.3.3 PTCM

Is the pedophile trend characterization module as shown in Figure 4. Once all the information has been collected, processed and implemented in BotHook, the system in response to each chat session performs a pedophile secondary classification (SPTC). When the system collects and catalogs in the PPTC and SPTC classifiers (primary and secondary pedophile classification), this information passes to an analysis in this module. This report is sent to the technical and psychological area of the project for a final validation.

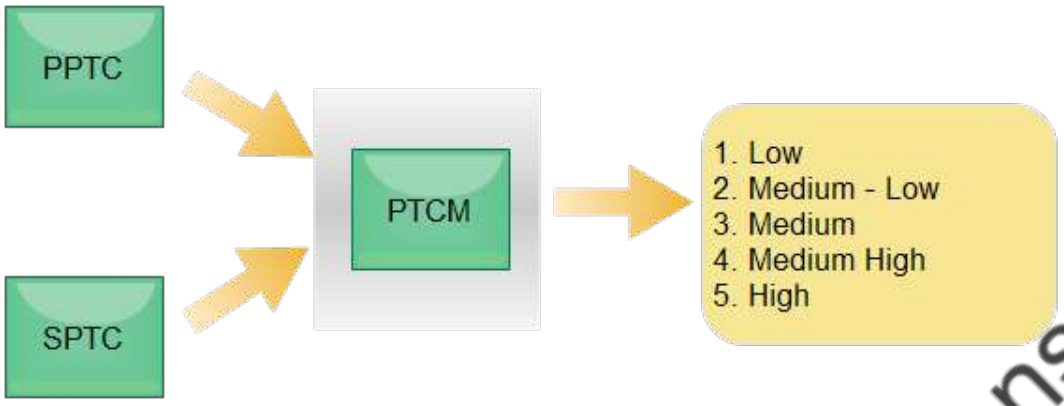


Figure 2.4: PTCM module

2.4 ANALYSIS AND DISCUSSION

The main contribution of BotHook is the collection of data through the sub module DAS. In this sub module, it is possible to classify potential computer criminals, and identify different characteristics of attacks to messaging systems where there is interaction with adolescents; determining their tools, attack tactics and relevant information that characterizes them. BotHook handles four edges that allow to maximize the credibility of this conceptual platform against a potential attacker and minimize the subjectivity in the classification.

The first edge consists of a set of AI techniques such as the use of neural networks (Seq2Seq model), classifiers and natural language programming, which already exist in the literature but in an isolated way. The platform searches to integrate these techniques, as part of the reproducibility of the scientific method applied in the project, with the aim of having a high percentage of success, measured by the teacher's (possible attackers) continuity in the tutorials with the system. The second edge is the creation of knowledge bases in conjunction with the psychological area, in order to obtain the largest number of questions-answers from the established student group. The knowledge bases available in proposal [6] can not be part of the project, because it establishes a Mexican slang that differs from Ecuadorian and does not clearly establish the age and gender that BotHook contemplates.

As a third edge, the BotHook project provides reliable information to the competent authority, since it is analyzed and validated by the psychological area once the system alerts a pedophile trend categorization in module PTCM. As a fourth edge, not contemplated in previous researches, the selection of the simulated age of the system and the subject (English) is justified. Hendley et al. in their research [11] describe that the English language is the

most popular among subjects such as math, science or technology among students aged between 11 and 13 years. Under this criteria, the selected subject is the English language, which acts as a means of attraction towards cyberpedophiles, without neglecting that it is considered a universal language.

McDonald-Brow et al. [20] describe that only 29 % of children between 9 and 10 years old have a profile on social networks and this increase to 59 % in the case of adolescents between 11 and 12 years old. In other study, Rich M. [12] examines the level of understanding of sexual content in digital media, determining that preadolescents and adolescents have a good understanding of sexual insinuations, but their comprehension at age of 12 years is significantly lower than adolescents between 14 and 16 years. Safekids.com does a categorization of Internet security by age and describes that in the age of 10 to 12 years, many children want to experience their independence using the Internet as a knowledge tool.

Considering the studies previously mentioned, we establish that the simulated age of BotHook must be 12 years, which represents the first year of school. It is at this age where children experience with more knowledge on the Internet, but with a lower level of understanding of sexual content. Under all the arguments described above it is concluded that the system provides a real need and a real scenario as a hook proposal in front of potential cybercriminals.

2.5 CONCLUSION

The diversity of proposals for modification, improvement or repowering of the chatbot based on ALICE, differs depending on the application that is wanted to obtain with the platform. It is important to emphasize that the research trend is to obtain better results of conversational response between the system and the user, leaving open a gap of study. Cybersecurity can and should rely on this platform to counteract cyberattacks such as grooming or cyberpedophilia in alignment with other sciences in a cross-sectional way. As future work, an experimental platform is proposed based on the conceptual platform presented in this research, BotHook. With the experimentation we will be able to analyze the computational cost that the above mentioned techniques demand. It is important to remark that as long as a system has a shorter response time, it will increase the chances of success of future projects in relation with the communication between a chatbot and an attacker.

REFERENCES

[1] Dante Contreras and Paulina Sepúlveda. Effect of lengthening the school day on mother's labor supply. *The World Bank Economic Review*, 2016.

[2] Eduardo Bieber Viola. Perfil psicosocial del pedófilo. In *El Perfilador*, pages 160–174. Universidad de la Rioja, 2012.

[3] Shweta Macwan et al. Web mining against pedophilia. *International Education and Research Journal*, 3(5), 2017.

[4] Charles Thorpe. The pornography of information. In *Necroculture*, pages 153–204. Springer, 2016.

[5] REGNER Sabillon, JEIMY Cano, VICTOR Cavaller, and JORDI Serra. Cybercrime and cybercriminals: A comprehensive study. *International Journal of Computer Networks and Communications Security*, 4:165–176, 2016.

[6] Ángel Callejas-Rodríguez, Esaú Villatoro-Tello, Ivan Meza, and Gabriela Ramírez-de-la Rosa. From dialogue corpora to dialogue systems: Generating a chatbot with teenager personality for preventing cyber-pedophilia. In *International Conference on Text, Speech, and Dialogue*, pages 531–539. Springer, 2016.

[7] Carlos Laorden, Patxi Galán-García, Igor Santos, Borja Sanz, Javier Nieves, Pablo G Bringas, and Jose María Gómez Hidalgo. Negobot: Detecting paedophile activity with a conversational agent based on game theory. *Logic Journal of IGPL*, 23(1):17–30, 2014.

[8] Majid Yar. Oxford research encyclopedia of criminology. 2017.

[9] Jiwei Li, Will Monroe, Alan Ritter, Michel Galley, Jianfeng Gao, and Dan Jurafsky. Deep reinforcement learning for dialogue generation. In *Proceedings of the 2016 Conference*

on Empirical Methods in Natural Language Processing, Austin, Texas, November 1-5, page 1192–1202, 2016.

[10] S Reshmi and Kannan Balakrishnan. Implementation of an inquisitive chatbot for database supported knowledge bases. *Sādhanā*, 41(10):1173–1178, 2016.

[11] David Hendley, Sian Stables, and Andrew Stables. Pupils' subject preferences at key stage 3 in south wales. *Educational studies*, 22(2):177–186, 1996.

[12] Michael Rich. Virtual sexuality: the influence of entertainment media on sexual attitudes and behavior. In *Managing the media monster: The influence of media (from television to text messages) on teen sexual behavior and attitudes, Washington*, pages 18–38. National Campaign to Prevent Teen and Unplanned Pregnancy, 2008.

3 HOW DOES GROOMING FIT INTO SOCIAL ENGINEERING?

Patricio Zambrano[1], Jenny Torres[1], Pamela Flores[1]

[1]Faculty of System Engineering, Escuela Politécnica Nacional, Quito, Ecuador

Email: {patricio.zambrano, jenny.torres, pamela.flores}@epn.edu.ec

3.1 ABSTRACT

In this research, we propose to formally include the technique of grooming within the processes of Social Engineering, validating its insertion with the phases of Social Engineering designed by Kevin Mitnick. Scientific evidence has shown that grooming is the technique applied by child sexual harassers and research in this field has generated significant contributions in this study. Nevertheless, researchers do not interrelate the contributions generated by grooming and Social Engineering, which allows us to infer that the studies are isolated and incomplete and must be consolidated with complementary scientific evidence.

KEY WORDS: Cyberpedophile, Grooming and Social Engineering Cycle.

3.2 INTRODUCTION

Nowadays, social engineering attacks are more common, specific and sophisticated. Cyber attackers take advantage of social behavior and norms such as reciprocity or social subtleties to gain access to information through human vulnerability. In the same way, they take advantage of the fact that most people never expect to be victims of social engineering, nevertheless, they are harmed by this type of attack, delivering unconsciously sensitive information. Some social engineering attacks have proven to be extremely expensive. In the

United Kingdom, it is estimated that crimes related to identity theft cost around 1.2 billion pounds in 2009. Phishing losses were around 23.2 million pounds in 2005 [1]. In 2004, the Department of Justice of USA concluded that one in three people could become a victim of social engineering during their lifetime [2].

This is a turning point where new questions are generated. Is only personal data the goal of social engineers? What types of techniques have been studied against human vulnerability? Are only professional people vulnerable to social engineering techniques? To answer these questions, we evaluated the types of Internet users and found that: in USA, two-thirds of households with children use the Internet. From these, 84 % of children and 97 % of young people between 12 and 18 are permanently connected due to the massive use of social networks [3]. This alarming data leads to a more detailed analysis of the different attacks to which this group of users would be vulnerable. According to the Child Protection and Online Protection Agency, online child harassment, known as grooming, was the most reported crime in the United Kingdom between 2009 and 2010 [4] [5]. Grooming is considered a process of preparation to approach, persuade and compromise the privacy and intimacy of a child or adolescent. After a more detailed analysis in the literature, we determined that grooming is not considered a social engineering technique, despite the efforts made to automatically determine the syntactic and semantic behavioral patterns of pedophiles online, based on the analysis of chats and P2P networks, when applying grooming as a technique of harassment [6] [7]. Krombholz et al. proposes a taxonomy of social engineering that cover users, channels and techniques [8]. Nevertheless, grooming is not included as a social engineering techniques.

In this paper, we propose an extension of this taxonomy to include the grooming as part of the social engineering processes. Our main objective is to expand the field of social engineering based on studies related to grooming [6] [7] thus creating greater knowledge that can be used as a tool to teach and as a framework to develop a holistic protection strategy.

The rest of this document has been organized as follows. Section II describes the different criteria and concepts of social engineering and grooming. Section III highlights the research methodology based on related studies. Section IV shows the interrelationship between social engineering and grooming. Section V and VI conclude the study by presenting a discussion of the results obtained and conclusions respectively.

3.3 BACKGROUND

3.3.1 Social Engineering

The identification and classification of threats are the basis for building defense mechanisms. However, humans are considered the weakest link in information security due to their susceptibility to different manipulation techniques. [1]. Concepts associated with "social engineering", define it as a dark art in computer science where, the use of deception is the main tool to induce a person to divulge private information involuntarily and, in this way, attackers gain access to computer systems [2]. The techniques employed by a social engineer include persuasion, coercion, urgency, authority, supplanting or requesting assistance, among others; taking advantage of human weakness [9, 10, 11]. In the case of social networks, it is very common for users to downplay the importance of the security of their information, as they rely on the protection mechanisms of the companies that run the social networking sites. Users tend to believe that popular companies like Facebook and Twitter will not allow anyone to exploit their information. However, instead of using technical means to exploit the user, social engineers use deception techniques to convince users to accept an attack. Kevin Mitnick said it's much easier to trick someone to present his credentials than using sophisticated piracy maneuvers [9]. The field of social engineering is still in its early stages of standardization with formal concepts, frameworks and work templates [12]. The art of influencing people to divulge sensitive information is known as social engineering and the process of doing it, is known as social engineering attack. Nowadays, there are several definitions of social engineering and several different models of social engineering attacks [1]. Many scientists strive in their studies to associate social engineering with non-technicality, while others include technical attacks that do not imply an important role for traditional social engineering from human to human. Academics such as Krombholz et al. [8], have proposed a theoretical / technical taxonomic analysis including technicality, as can be seen in Figure 3.1.

This study is based on two previous proposals [13, 14]. Our study considers taking this taxonomy as a reference point, for its theoretical and technical support, as well as the four phases proposed by Mitnick [1] to determine if an attack is associated with social engineering: 1) information gathering, 2) development of relationship, 3) exploitation of relationship; and 4) execution to achieve objective.

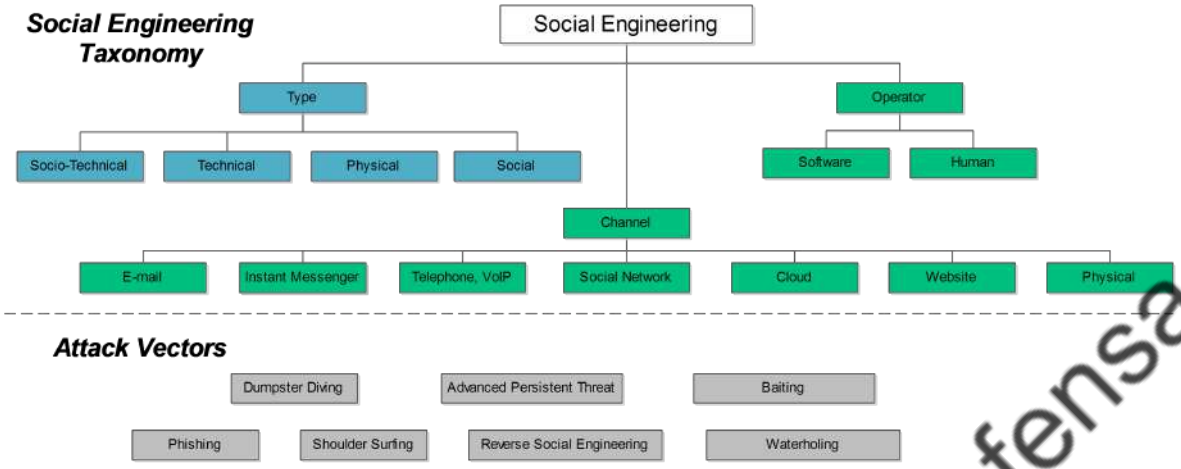


Figure 3.1: Krombholz et al. Social Engineering Taxonomy

3.3.2 Grooming

Grooming has been studied for more than a decade and these studies have generated significant contributions to society [6] [7] [15]. In the scientific field, grooming has been conceptualized as a procedural technique used by cybernetic attackers, who in some cases are pedophiles or pederasts. On the other hand, it is also considered as an operational concept, through which an aggressor applies affinity search strategies, while acquiring information and sexually desensitizing the victims to develop relationships that result in the satisfaction of the attacker's needs [6]. Through different studies, it is evident that the grooming can be applied for several years, in order to prepare the victims and guarantee cooperation, thus minimizing the risk of exposure by the attacker to the victim. In some cases, it is also considered the preparation of relatives close to the victims to create an atmosphere of acceptance and normalization of the attack. Perverted-justice is an online tool, which aims to eradicate online predators [6]. Since its inception, they planned to publish chats of real predators, thus exposing them. This was the starting point where scientists began to study and analyze the text chains published by this web portal, thus determining psychological and technical behavioral traits when applying grooming as a preparation for victims. One of the biggest challenges evidenced in the studies of chat chains is phonetics and phonology, where the fields of study of morphology, syntax, semantics, pragmatics and discourse are derived. Within these studies it has been determined that online pedophiles tend to seduce their victim through attention, affection, kindness and even gifts. In a survey applied to 437 schoolchildren between 11 and 13 years, they use habitually chat rooms. Fifty-nine percent of the participants say that they have regularly participated in chats with people over the

Internet. The 24 % of people who chatted online admitted having delivered some type of personal information. These include the house telephone number, the mobile phone number and the address of the house. The most alarming fact in this study was that 37 children admitted to decide to meet the person they were chatting with [16]. In our analysis and technical questioning of the proposals of the grooming stages, see Figure 3.2, we observe that the terms are very general or ambiguous. Olson, Welner, O'Connell, evidence and justify different stages of grooming, these being the most cited by related research [17][5][4]. Based on that, we use this set of technical criteria to propose 4 stages that show a level of access to the victims:

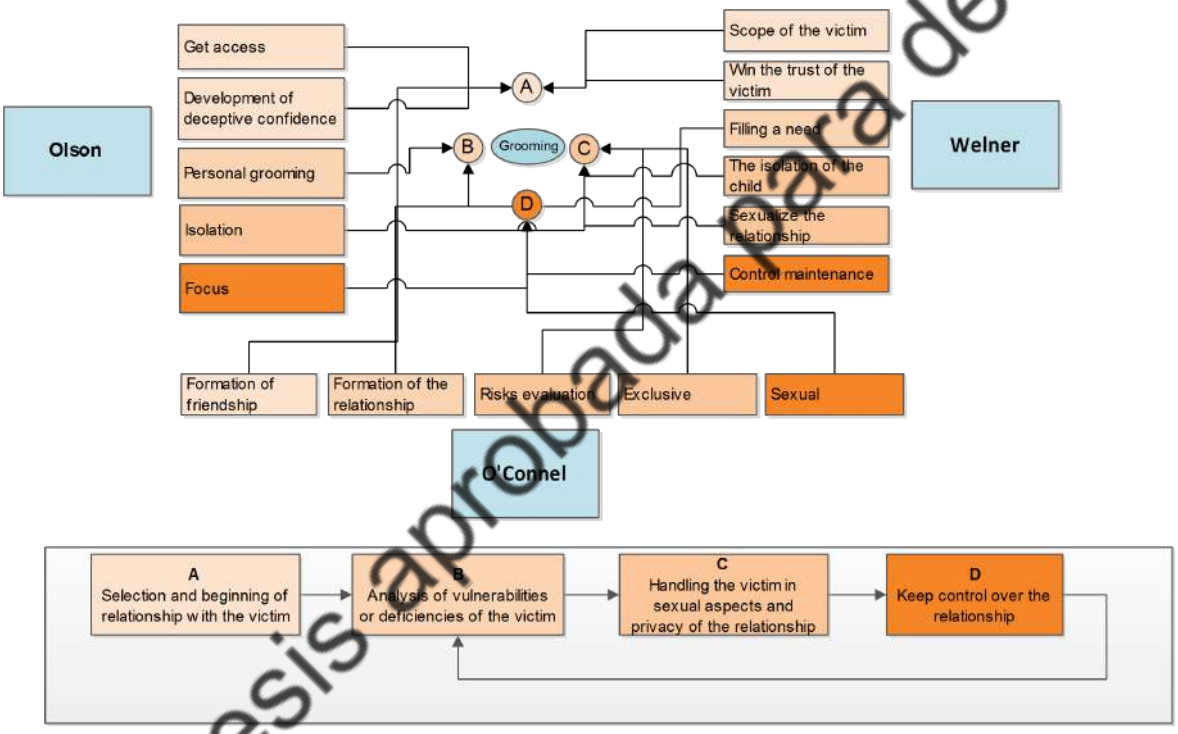


Figure 3.2: Grooming stages

A) Selection and beginning of relationship with the victim. In every attack, the perpetrator analyzes and selects his objective, for our case a child or adolescent who provides the ideal conditions (ethnicity, deprivation, lack of attention, etc.) to start a relationship of trust. B) Analysis of vulnerabilities or deficiencies of the victim. Once the relationship starts, the attacker will proceed to analyze their vulnerabilities and strategically exploit each of them, for our case, the attacker will proceed to fill spaces where the victims feel isolated, lacking attention or suffer from some type of economic need. C) Handling the victim in sexual aspects and privacy of the relationship. The victim will be involved in the sexual field once the attacker has achieved a space of marked confidence. This process will require time and confidentiality in the exchange of information since the attacker will sometimes share pho-

tos of their own sexual content or that of other victims to desensitize them and to establish confidentiality rules so that the victim does not disclose their conversations. D) Keep control over the relationship. In this stage the attacker has become a very important person for the victim and in his eagerness to preserve the domain of the relationship he will apply again the techniques of vulnerability.

3.4 METHODOLOGY

The inductivist approach is the general basis of this research, since it draws general conclusions, starting from particular observational statements. This method is characterized by having four basic stages: observation and recording of all the facts, analysis and classification of the facts, inductive derivation of a generalization from the facts; and contrast [2]. *i) Observation and registration*.- We begin by observing the problem, online pedophilia in contrast to social engineering, and the need to generalize a model that encompasses them by recording all the scientific documentation that deals with these topics. *ii) Analysis and classification*.- To make this generalization, the study was based on existing theoretical knowledge.. The observations were made on studies of attacks of real online pedophiles that use grooming as an attack technique, and the analysis of studies and taxonomies related to social engineering [8]. The study of the literature showed that the topics under discussion are part of present and future research. Various characteristics that describe them have been investigated, analyzed and published, except for their interrelation. [8] [9] [18]. *iii) Inductive derivation of a generalization*.- To better understand cyber-pedophilia and the use of grooming as an attack technique, we endeavor to demonstrate well-documented and representative studies of grooming and cyber-pedophilia [6] [17] [15]. A particularity of the investigations related to pedophilia is that they analyze the texts of real conversations of pedophiles published in the preverted-justice website. These studies are very diverse and range from neuro linguistic programming (NLP), textmining, mathematical models to a wide gamma of artificial intelligence techniques to determine a profile or pattern of pedophile behavior. These studies demarcate a considerable analysis of pedophilia and refer to grooming as the technique used to access the victims. Under this criterion, it is considerable to infer that grooming is part of the generalization of the concept of social engineering. And finally *iv) Testing*.- It should be noted that grooming, in the investigations that were part of our study, did not infer that it is part of social engineering. Under this premise, a more detailed study of social engineering was acomplished, considering vectors of attacks, applied techniques, taxonomies and communication channels.

3.5 RESULTS

As stated above, for an attack to be classified as a Social Engineering attack, it must fulfill Mitnick's 4 phases: 1) Information gathering, 2) Development of relationship, 3) Exploitation of relationship, and 4) Execution to achieve objective. The grooming as an attack technique has shown certain characteristics that can be clearly classified within two Mitnick phases: development of relationship; and exploitation of relationship.

3.5.1 Information gathering

This phase consists on collecting as much information as possible about the target. This information is used in the next phases and is of vital importance for the quality of the result in cases of targeted attacks. In the study about grooming, we consider it relevant to understand the psychological-technical profile of the attacker, since the type of attack and the information gathered depend on it.

3.5.1.1 Types of attackers

In the literature, we can identify different criteria of the attackers since they are treated as child sexual abusers or pedophiles. The latter is sometimes considered a phenomenon under constant study, however, in other cases, it is treated as a disorder of sexual preference "pedophile" and the consequence of their behavior is named as adolescent sexual abuse (CSA) [19] [20]. Pedophilia is defined by the diagnostic manual of the World Health Organization (WHO) and the International Classification of Diseases (ICD-10) as: "a disorder of sexual preference". It should be mentioned that ICD-10 indicates that pedophile behaviors are very heterogeneous. These behaviors can be harmless, even those that reach levels of child sexual abuse or pedophilia [21]. However, there is another type of attacker. This one does not have a disorder of sexual preference towards children but sees in child pornography a profitable business type and can use the grooming to obtain child pornographic material, for its subsequent sale and distribution. In the scientific field, the pedophile is considered the only user who uses grooming to persuade his victim and commit an act of rape. Under this consideration, our investigation raises three types of criminals and their interests in the field of pedophilia: the cyber-pedophile, cyber-pederast or child sexual abuse, and the cyber-offender. After the analysis of the potential attackers it is possible to infer the type of information that these will gather in the process of attack: files with pornographic content, personal information, address of domicile and places of frequent visit, etc., being their po-

tential targets, the personal use and/or commercialization of pornographic material obtained and in the worst scenario, the violation.

3.5.2 Development of relationship

In this phase, related researches describe characteristics of development of relationship without linking them to social engineering. It is noted that, to locate children or people with access to a child, pedophiles use social networking sites, blogs, chat rooms for children, IM, email, discussion forums and online children's play websites, to start a relationship. The online games of PC, Xbox, PlayStation and Wii, are spaces where dangerous relationships are developed between a child and a delinquent [22]. An important aspect of development of relationship is mentioned in [23] where pedophiles seduce their young prey through attention, affection, kindness, and even gifts. Two additional elements to the relationship and trust are: simulation of child behavior (slang), it is generally used by predators to copy their linguistic style and the second aspect is the sexual language, where predators gradually change a normal conversation to a sexual one, starting with more ordinary compliments [20]. The development of the relationship and trust by stages has been evaluated in [17], from the point of view of the development of deceptive trust that consists of building a relationship of trust with the child. This stage allows the predator to build a common ground with the victim.

3.5.3 Exploitation of relationship

The exploitation of the relationship is another feature considered by Mitcknic in [18], given that, it demonstrates the success of this exploitation by achieving the exchange of pornographic content files with underages. Harms mentions that an aggressor applies affinity search strategies, while acquiring information and sexually desensitizing its specific victims to develop relationships that lead to the satisfaction of needs [6]. After the previous stage, the predators maintain a fixed discourse, in which they are not willing to depart from the sexual conversation [20]. The elimination of communication records and the transfer of responsibility to the victim are two traits marked in this investigation. [17] represents the attack by cycles and in one of them it, shows the isolation of the victim from his friends and relatives once the relationship has intensified and, consequently the predator seeks to physically approach the minor. In this stage, the predator requests information such as, the child's and father's schedules, and the child's location.

3.5.4 Execution to achieve objective

The first use of the information gathered is the sale and distribution of pornographic material. The profile of the associated attacker, to this process, is the cybercriminal that based on aggression, which in some cases is agreed with family members, obtains audio or visual material. The private use of pornographic material is for the exclusive use of cyber-pedophiles who do not intend to commit rape, however, within the development of relationship, they exchange pornographic material to desensitize their victims and even use aggression to persuade them to keep the relationship secret. The pedophile or pederast, who in some cases is considered a child sexual abuse is the most dangerous because within the fields of exploitation of relationship and the development of relationship is the one that uses more mechanisms to get information of his victim with the aim of sexually abusing her.

3.6 DISCUSSION

The correlation between the phases of Mitnick and the processes associated with grooming, begin with the definition of the types of attackers and their objectives (information collection), because they will be responsible for collecting the necessary information to proceed with the next phase. In the development of the relationship, Figure 3.3., the attackers apply tactics such as persuasion, offline meetings, alternative communications, etc. In this way, if they achieve their objectives, they move on to the next phase of “exploitation of the relationshi”, where the attackers achieve the isolation of their victims, sexual desensitization of these, obtain the information they request, etc. Finally, they achieve their objectives, such as the sale and distribution of pornographic material, private use of the material and the execution of pedophilia. The grooming by its nature must be located within the attack vectors of the proposed taxonomy, as shown in Figure 3.4, establishing the relationship with the various channels that allow access to its victims: email, instant messaging, social networks, websites and in some cases physical encounters. The social aspect is established as the type of attack and the human aspect as operator.

The adaptation of grooming to the taxonomy of Krombholz et al. can be used in the scientific field in different ways. It can be used to educate about social engineering based on real experiential knowledge. As the taxonomy covers the actors and activities related to the field of social engineering (victims, attackers and the protective organization), it offers a holistic and comprehensible vision. It also facilitates a deeper understanding of the grooming process, and perhaps most importantly, offers an easy way to understand how to develop a

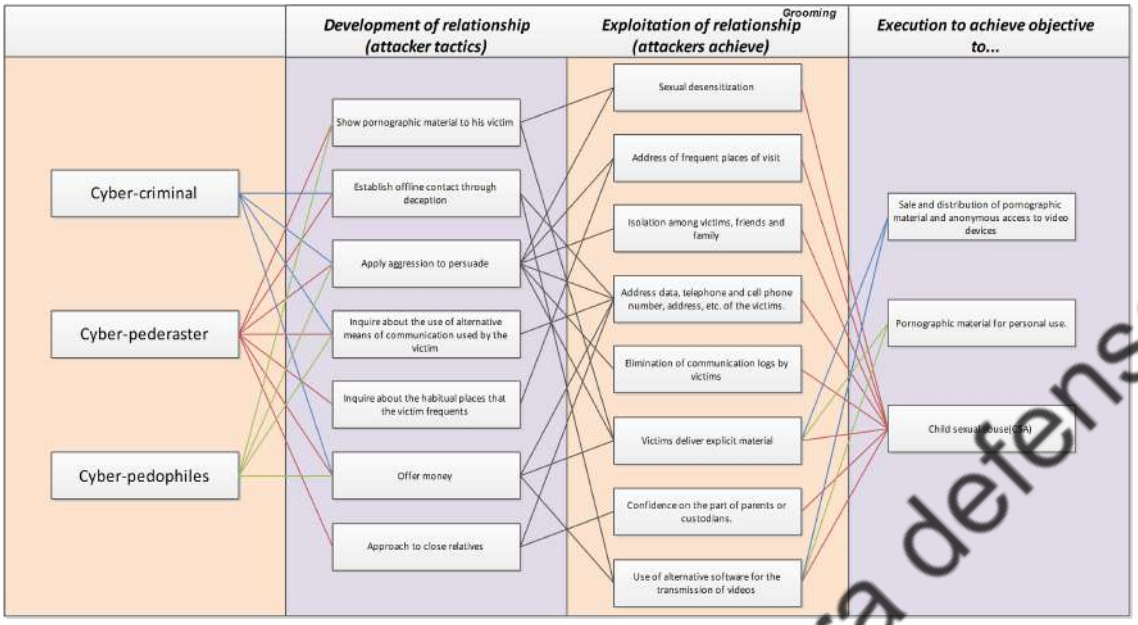


Figure 3.3: Summary of Mitnick phases and grooming

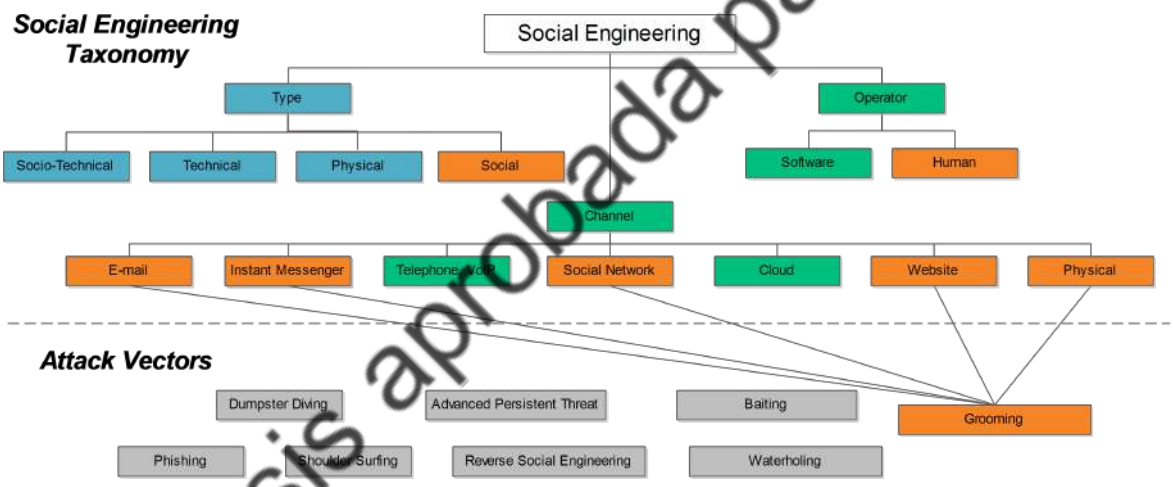


Figure 3.4: Proposed taxonomy and grooming

protection strategy. From the academic point of view, this research presents a start point for security researchers who try to position themselves in various fields related to pedophilia. By complementing the taxonomy, from the computer science point of view, computer applications can be generated to facilitate the early detection of possible attacks on children and adolescents using computer learning.

3.7 CONCLUSIONS

Grooming as an access technique has shown in our research that it focuses on a very sensitive group of society, children and adolescents. It is possibly the most dangerous technique within security information since it not only attacks against material assets but against the

emotional and mental stability of people that could bring sequels for a lifetime. Computer science in its continuous contribution to society, is establishing new aspects of social engineering, for a better understanding and research of this phenomenon. We established the interrelationship between social engineering and grooming through the 4 phases of the Mitnick cycle. In our eagerness to position and encourage the study of grooming, it is considered that it should be part of a taxonomy already agreed among scientists, thus generating a significant contribution to the development of previously established knowledge.

REFERENCES

[1] Francois Mouton, Louise Leenen, and H. S. Venter. Social engineering attack examples, templates and scenarios. *Computers and Security*, 59:186–209, 2016.

[2] Marcus Nohlberg and Stewart Kowalski. The cycle of deception: a model of social engineering attacks, defenses and victims. In *Second International Symposium on Human Aspects of Information Security and Assurance (HAISA 2008), Plymouth, UK, 8-9 July 2008*, pages 1–11. University of Plymouth, 2008.

[3] Saba Mohammed and Edward Apeh. A model for social engineering awareness program for schools. In *Software, Knowledge, Information Management & Applications (SKIMA), 2016 10th International Conference on*, pages 392–397. IEEE, 2016.

[4] Fergyanto E. Gunawan, Livia Ashianti, Sevenpri Candra, and Benfano Soewito. Detecting online child grooming conversation. *Proceedings - 11th 2016 International Conference on Knowledge, Information and Creativity Support Systems, KICSS 2016*, 2017.

[5] Hady Pranoto, Fergyanto E. Gunawan, and Benfano Soewito. Logistic Models for Classifying Online Grooming Conversation. *Procedia Computer Science*, 59(Iccsci):357–365, 2015.

[6] Nick Pendar. Toward spotting the pedophile telling victim from predator in text chats. *ICSC 2007 International Conference on Semantic Computing*, (c):235–241, 2007.

[7] India McGhee, Jennifer Bayzick, April Kontostathis, Lynne Edwards, Alexandra McBride, and Emma Jakubowski. Learning to Identify Internet Sexual Predation. *International Journal of Electronic Commerce*, 15(3):103–122, 2011.

[8] Katharina Krombholz, Heidelinde Hobel, Markus Huber, and Edgar Weippl. Advanced social engineering attacks. *Journal of Information Security and Applications*, 22:113–122, 2015.

[9] Taimur Bakhshi. Social engineering: revisiting end-user awareness and susceptibility to classic attack vectors. In *Emerging Technologies (ICET), 2017 13th International Conference on*, pages 1–6. IEEE, 2017.

[10] Rüdiger Wirth and Jochen Hipp. Crisp-dm: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, volume 1, pages 29–39. Manchester, 2000.

[11] Colin Shearer. The crisp-dm model: the new blueprint for data mining. *Journal of data warehousing*, 5(4):13–22, 2000.

[12] Kevin D Mitnick, William L Simon, and S Wozniak. The art of deception: Controlling the human element of security. 2002. *Paperback ISBN 0-471-23712-4*, 2006.

[13] Koteswara Ivaturi and Lech Janczewski. A taxonomy for social engineering attacks. In *International Conference on Information Resources Management*. Centre for Information Technology, Organizations, and People, 2011.

[14] F Mohd Foozy, Rabiah Ahmad, M Abdollah, Robiah Yusof, and M Mas'ud. Generic taxonomy of social engineering attack. In *Malaysian Technical Universities International Conference on Engineering & Technology*, pages 527–533, 2011.

[15] Ryan C.W. Hall and Richard C.W. Hall. A profile of pedophilia: Definition, characteristics of offenders, recidivism, treatment outcomes, and forensic issues. *Mayo Clinic Proceedings*, 82(4):457–471, 2007.

[16] Katrinna MacFarlane and Violeta Holmes. Agent-mediated information exchange: Child safety online. *Proceedings - International Conference on Management and Service Science, MASS 2009*, 2009.

[17] A.E. Cano, M. Fernandez, and H. Alani. Detecting child grooming behaviour patterns on social media. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8851:412–427, 2014.

[18] Surbhi Gupta, Abhishek Singhal, and Akanksha Kapoor. A literature survey on social engineering attacks: Phishing attack. In *Computing, Communication and Automation (ICCCA), 2016 International Conference on*, pages 537–540. IEEE, 2016.

[19] Dasha Bogdanova, Paolo Rosso, and Thamar Solorio. Modelling Fixated Discourse in Chats with Cyberpedophiles. *Proceedings of the Workshop on Computational Approaches to Deception Detection*, pages 86–90, 2012.

[20] Dasha Bogdanova, Saint Petersburg, Paolo Rosso, and Thamar Solorio. On the Impact of Sentiment and Emotion Based Features in Detecting Online Sexual Predators. *Proceedings of the 3rd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, (July):110–118, 2012.

[21] Anna Vartapetiance and Lee Gillam. "Our Little Secret": pinpointing potential predators. *Security Informatics*, 3(1):1–19, 2014.

[22] Lyta Penna, Andrew Clark, and George Mohay. A framework for improved adolescent and child safety in MMOs. *Proceedings - 2010 International Conference on Advances in Social Network Analysis and Mining, ASONAM 2010*, pages 33–40, 2010.

[23] Siew Yong, Dale Lindskog, Ron Ruhl, and Pavol Zavarsky. Risk mitigation strategies for mobile Wi-Fi robot toys from online pedophiles. *Proceedings - 2011 IEEE International Conference on Privacy, Security, Risk and Trust and IEEE International Conference on Social Computing, PASSAT/SocialCom 2011*, pages 1220–1223, 2011.

4 TECHNICAL MAPPING OF THE GROOMING ANATOMY USING MACHINE LEARNING PARADIGMS: AN INFORMATION SECURITY APPROACH

Patricio Zambrano[1*], Jenny Torres[1], Luis Tello[2], Rubén Jácome[1], Marco Benalcázar[1], Roberto Andrade[1], Walter Fuertes[3]

[1]Department of Informatics and Computer Science, Escuela Politécnica Nacional, Quito, Ecuador,

[2]Faculty of Engineering, Universidad Nacional de Chimborazo, Ecuador,

[3]Faculty of System Engineering, Universidad de las Fuerzas Armada, Sangolquí, Ecuador

4.1 ABSTRACT

In the field of information security, there are several areas of study that are under development. Social engineering is one of them that addresses the multidisciplinary challenges of cyber security. Nowadays, the attacks associated with social engineering are diverse, including the so-called Advanced Persistent Threats (APTs). These have been the subject of numerous investigations; however, cybernetic attacks of similar nature as grooming have been excluded from these studies. In the last decade, various efforts have been made to understand the structure and approach of grooming from the field of computer science with the use of computational learning algorithms. Nevertheless, these studies are not aligned with information security. In this work, the study of grooming is formalized as a social engineering attack, contrasting its stages or phases with life cycles associated with APTs. To achieve this goal, we use a database of real cyber-pedophile chats; this information was refined and the Latent Dirichlet Allocation (LDA) topic modeling was applied to determine the stages of the attack. Once the number of stages was determined, we proceed to give them a linguistic context, and with the use of machine learning, a linear model was trained to obtain 97.6 % of training accuracy. With these results, it was determined that the study of

grooming could support research associated with social engineering and contribute to new fields of information security.

KEY WORDS: Cyber-pedophile, Pedophile, Grooming, Pattern Behavior, APT, Social Engineering.

4.2 INTRODUCTION

Recently, attacks on the privacy of children and adolescents through technological means have increased considerably. Investigations related to this social problem addressed different topics such as: the study and analysis of children's images in P2P networks, planning of security models in audio-visual devices for child control with access to the Internet, study of vulnerabilities in online video games, development of communication bots for the detection of potential attackers, forensic tools and analysis of pedophile behavior within instant messages [1, 2, 3, 4, 5, 6]. Several scientific proposals that have studied the behavior of online attackers stand out in the study of instant messaging. Researchers in this field determined that the most common technique applied by attackers is grooming.

This technique is characterized by using deceptive linguistic expressions to create environments of false friendship and trust, thus seducing victims to manipulate and gaining control over them. Some authors, supported by previous research and psychological studies apply techniques of text mining and machine learning to determine the nature and different levels of danger of this attack [7, 8, 9, 10]. However, it has been shown that the results of the research related to the study of grooming, contemplate different lines of research, are not conclusive and support other relevant studies [11, 12, 13, 6].

Online pedophilia and grooming have been studied for over a decade [1, 2, 14, 15, 3]. In the scientific field, grooming has been conceptualized as a procedural technique used by cybernetic attackers [1]. On the other hand, it is also regarded as an operational concept, whereby an attacker applies search strategies affinity, while acquires information and sexually desensitizes victims to develop relationships that lead to the satisfaction of the needs of the offender or attacker [2]. The main motivations of this research are to delimit the technical anatomy of grooming, justifying its relevance and support for future investigations of relevance. These attacks are also known as social engineering semantic attacks and are considered pervasive threats to computer systems, communication, and privacy [16].

In this study, as the first phase, we propose a technological alternative that allows the grooming life cycle to be determined through stations or phases named *topics* with the use of the Latent Dirichlet Allocation (LDA) generative probabilistic framework, belonging to the field of topic modeling within text mining. LDA allows modeling the topic structure of documents and other discrete data collections, where each document is generated as a mix of topics. In this way, LDA assigns a topic to a set of words contained in each document [17]. Then, two experiments are proposed. In *experiment 1*, several topics are determined according to the characteristics of the pre-processed data. To obtain the data and its processing, the recommendations of the CRISP-DM methodology [18] were followed. After determining an optimal number of stations, we proceeded to give them a logical context through *experiment 2*, which uses studies related to linguistics and communicational intentions to order the topics determined by LDA. Within this ordering, several proposals of life cycles of Advanced Persistent Threats (APT) with the topics were related, thus determining the life cycle of the grooming.

The main contributions of this study are summarized as follows

- A psychological and technical profile of the type of attacker associated with online pedophilia is presented; Uncovering a technical psychological profile of online attackers uncovers characteristics and behavioral patterns that can be used to analyze and execute technical actions to proactively protect future victims.
- Grooming as a vector of attack within social engineering and information security is positioned; this will allow supporting investigations related to determining patterns of malicious behavior online;
- Through the modeling of topics, different stages or seasons of a life cycle of an attack associated with social engineering is determined;
- Application of a linear machine learning algorithm to classify texts binding to the study area.

The article proceeds as follows. Section 4.3 introduces some definitions about grooming, its stages, and tools to detect it. Section 4.4 establishes the physicological/technical profile of a cyber-pedophile based on the use of technological resources. Section 4.5 presents the related work on the topic. Section 4.6 details the methodology we follow, defines the research questions, and introduces the experimental approach. Section 4.7 develops the

experiments carried out. Section 4.8 presents the answers to the research questions based on the results of the experimental phase. Finally, Section 6.7 draws the conclusions and present the future work.

4.3 WHAT IS GROOMING?

With the advance and use of communication technologies, the evolution of pedophile attacks in cyberspace and their strategies in approaching their potential victims has been evidenced. Within these strategies, there is the attack known as *grooming* that is used to create deceptive trust relationships between victims and attackers. The grooming has been considered as a case study, for more than a decade. From this, related research has generated significant contributions to society [1, 2, 14, 15, 3].

In the scientific field, grooming has been conceptualized as a procedural technique, criminal activity, or operational concept used by cybernetic attackers, who in some cases have a disorder of sexual preference for children or adolescents. In the development of false friendship, the attacker applies strategies to determine affinities, tastes or activities of interest to the victim, thus developing a relationship of trust where the main objective is sexual desensitization, giving rise to the satisfaction of the needs of the attacker; as the sexual act [2, 19, 1]. As a computer attack technique, grooming can be applied for very long periods, in order to guarantee the cooperation of its victims and minimizing the risk of exposure. Another aspect to be considered within the technique of grooming includes the preparation of relatives close to the victims to create an atmosphere of acceptance and normalization of a potential attack [1, 20, 21].

The study of grooming is comparable to the study and analysis of modern and contemporary computer attacks. The development of proactive measures and the advancement of investigations are limited by several aspects, such as access to databases of pedophile content, victims and relatives and means of communication, such as the Internet.

The online activist Perverted-justice (PJ) Foundation has collaborated extensively with the study of pedophile communications by continuously publishing actual conversations on its website. The primary purpose of this foundation is to eradicate online attackers by exposing the conversations and their actors [2]. With this background, several scientists began to study and analyze the text strings published by this web portal, thus determining psycholo-

gical and technical behavioral traits when applying grooming in preparation for victims. One of the most significant challenges evidenced in the studies of chat chains is phonetics and phonology, which reveal the fields of study of morphology, syntax, semantics, pragmatics, and discourse.

Studies have determined that online pedophiles tend to seduce their victims through attention, affection, kindness, and even gifts through the use of information technologies [20]. Such is the case that in a survey applied to 437 schoolchildren between the ages of 11 and 13 years, it was determined that the use of the Internet and the chat communication protocol were part of their regular habits. 59 % of the participants accepted to have participated regularly in chats with people through the Internet. 24 % of people who chatted online admitted having delivered some personal information. These include the home phone number, the mobile phone number, and the home address. The most alarming fact in this study was when 37 children admitted to making arrangements to meet the person they were chatting with [22].

4.3.1 Stages of Grooming

When talking about grooming, an essential strategy for an attacker is the **sexual desensitization**. Kong et al. in [23] consider it as a common strategy that offenders use for a child to access the sexual encounter. This sexual desensitization tends to occur gradually. Usual physical or emotional contact, such as bathing, cuddling, or tickling, can eventually become sexual contact and then possibly more intrusive forms of sexual abuse. In fact, almost two-thirds of the children in the study indicated that at first, the genital contact seemed accidental. It should be noted, however, that some of the victims pointed out that the change from usual physical contact to sexual abuse was abrupt and, therefore, the period of gradual sexual desensitization was small or nonexistent. Attackers also endorsed the use of tactical sexual desensitization. In this study, it was evidenced that around a quarter of the attackers who care for their victims admitted having used these grooming techniques. Besides, almost a third of the attackers admitted to having asked the child for help with something, such as undressing. Almost half admitted having talked about sex with the child or having “accidentally” touched the child. Attackers also admitted to using pornographic videos and magazines to desensitize the child to sex.

It is worth noting that the use of pornography in children with sexual insensitivity is more

common among male victims than among women. The offender may tell the child that he is teaching him/her sexual education using photographic resources and the body of the victim. The research also emphasizes that attackers gradually increase physical contact. For example, the offender can start fighting, kissing, massaging, or curling up the child, all while evaluating the child's reaction to touching. If the child feels uncomfortable and asks the offender to stop, he may stop for a moment and then gradually increase the contact. The use of games, for example, *Red Light-Green Light* is also used for this purpose. In this situation, the offender may begin to touch the child's leg until the child protests. Other conventional techniques that the offender can use to desensitize the child is to "accidentally" show his naked body to the child, making sexual comments about the child's body or clothes, or telling him about previous sexual encounters that he or she has had.

Rutgaizer et al. [24], justify and assure that in the scientific field, there have been few investigations to understand the behavior patterns of sexual attackers in the different stages of online child harassment. In these stages, we observe the development of deceptive confidence, preparation, and the search for a physical encounter. In this research, characterizing the stages becomes a highly critical aspect, since most of the sexually abused children have been forced to accept physical encounters with the sexual attacker voluntarily. This suggests that understanding the different strategies that an attacker uses to manipulate children's behavior could help to educate them if they are exposed to these types of situations where their integrity is at risk. The research is developed based on Olson's Luring Communication Theory (LCT) [25], where once an attacker has had access to a child, the stages are:

- **Development of deceptive confidence:** consisting of developing a relationship of trust with the child. At this stage, the attacker exchanges personal information such as age, tastes, and distraction activities. This stage allows the attacker to build a shared communication space with its victim. Once a relationship of trust is established, the attacker proceeds to the next stage;
- **Grooming stage:** in this stage, the attacker triggers the sexual curiosity using sexual terms; it is at this moment where the attacker can prepare and catch communicatively the child in an online sexual behavior;
- **Entrapment cycle:** as the grooming process intensifies, the attacker oversees manipulating the victim so that she isolates herself from her friends and family, which

promotes and increases the trust relationship between attacker and victim;

- **Physically approach:** in this final stage, the attacker seeks to approach to the minor. The attacker requests information related to the child's location and schedules and their family members.

In [26], Hofman et al. describe 17 descriptors of the grooming process within six stages of work:

- **friendship:** the attacker tries to approach the child by determining similarities, tastes and activities in common, and on the other hand the attacker searches based on photographic evidence requested from the minor and alternative online methods that contrast the information with the child, to confirm that he is a child.
- **relationship:** the attacker and the child talk about the family, the school, the interest and the hobbies of the child in order to exploit them deceptively, making the child believe that they are in a relationship.
- **risk assessment:** in this stage, the author attempts to measure the level of threat and danger by talking to the child. He makes sure that the child is alone and that no one else is reading their conversations.
- **exclusivity:** the attacker tries to gain the child's full confidence. Frequently, the attacker introduces the concept of love and care in this stage.
- **sexual:** the attacker and the child talk about sexual activities and develop sexual fantasy.
- **conclusion:** at this stage, the attacker approaches the child to meet in person.

It is worth noting that the researchers describe that these grooming stages may or may not occur in the same sequence. The frequency, order, and extent of the occurrence of these stages may vary depending on the case.

4.3.2 Tools and Technologies to detect grooming

Figure 4.1 describes the diverse tools used to detect grooming. The numbers correspond to a sample of studies described in Table 4.1. These tools are classified into:

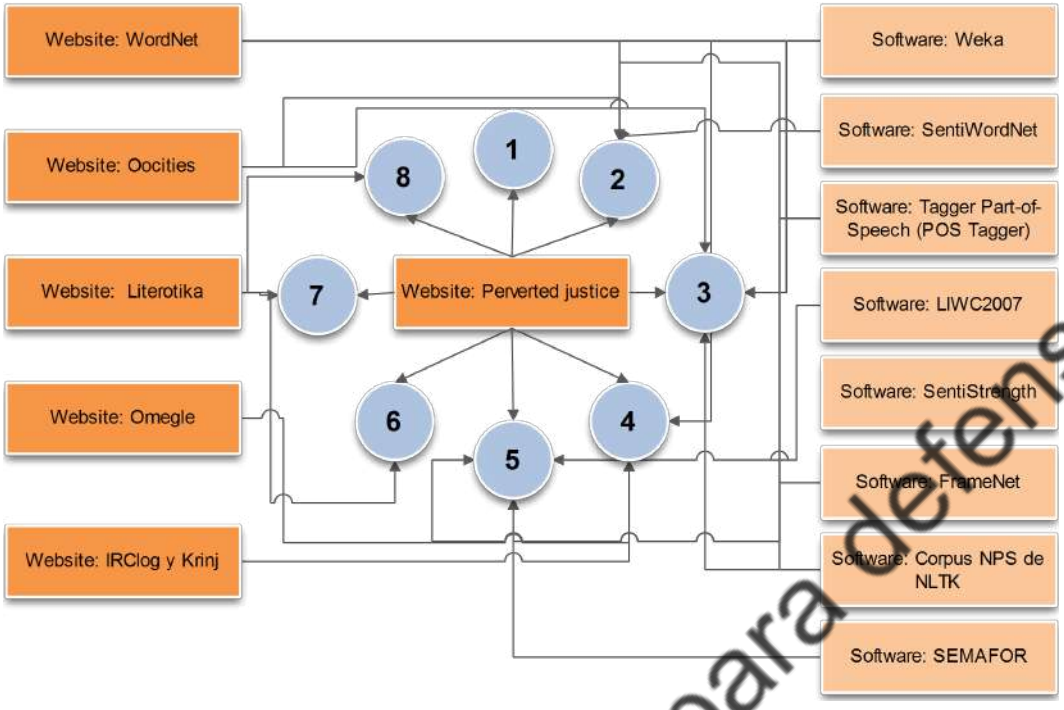


Figure 4.1: Graphic scheme that relates a group of numbered researches (more related to research). All researchers use data from Perverted Justice and process the information with a set of technological tools based on their objectives.

Table 4.1: Research Related To The Study Of Grooming

Item	Research	Reference
1	Toward Spotting the Pedophile Telling victim from attacker in text chats	[2]
2	On the Impact of Sentiment and Emotion Based Features in Detecting Online Sexual Attacker Dasha	[9]
3	modeling Fixated Discourse in Chats with Cyber-pedophiles	[27]
4	“Our Little Secret”: pinpointing potential attacker	[28]
5	Detecting Child Grooming Behaviour Patterns on Social Media	[14]
6	Exploring high-level features for detecting cyber-pedophilia	[29]
7	Logistic Models for Classifying Online Grooming Conversation	[19]
8	Detecting Online Child Grooming Conversation	[30]

❖ **Website**.- The investigations related to the field of pedophilia are based on the collection of real conversations of pedophiles for further analysis. PJ Foundation is the main provider of this information. In the analysis of the scientific proposals to detect grooming it could be determined that when creating artificial intelligence algorithms, these should have the minimum error rate, which is why the researchers contrasted

their models with chats of pedophile and non-pedophile sexual content. These chats were obtained from websites such as Oocities, IRClog, Krinj, among others.

- **Software**.- The use of specialized software has allowed researchers to extract positive and negative words to analyze the technical profile, behavioral patterns, applied discourse, sentiment analysis, semantic analysis of attackers among others. The use of software specialized in artificial intelligence allowed the researchers to execute and test some algorithms of automatic learning and data mining.

- **Corpus or Database**.- NPS Chat Corpus is a closed set of texts or information intended for scientific research and is part of the Natural Language Toolkit package (NLTK). NLTK is a natural language processing platform that allows researchers to build programs with human language data and thus generate predictions of conversations and behavioral patterns.

4.4 PSYCHOLOGICAL/TECHNICAL ATTACKER PROFILE

For the study of psychological profiles based on the use of technological resources such as the Internet, one of the pioneering investigations in the classification of types of attackers and victims is the one proposed by Pendar [2] (see Figure 4.2). In this study, two types of scenarios and their actors are considered. The first scenario is where Attacker/Other (1) are interrelated and the second Adult/Adult (2) where there is a consensual relationship. Three types of actors emerge from the first scenario: (a) Attacker/Victim where the victim is a minor. (b) Attacker/Pseudo-Victim in this case the victim is a volunteer posing as a child and (c) Attacker/Pseudo-Victim where the victim is an officer of the law pretending to be a child.

Ideally, to build a computer system that signals an interaction as suspicious, that is of type (1a), at least it is necessary to have access to representative samples of type 2 interactions also like (1a). However, this research indicates that chat service providers do not usually archive chats files for adults, and even if they did, they would not make those files available to the public. The accumulation of such data requires the informed consent of the participants. In addition, access to chat text files of type (1a) is also very difficult to achieve. Obtaining access to the data types (1c) is not without problems, since legal problems must be resolved in terms of privacy. Therefore, even a simple feasibility study for this type of research propo-

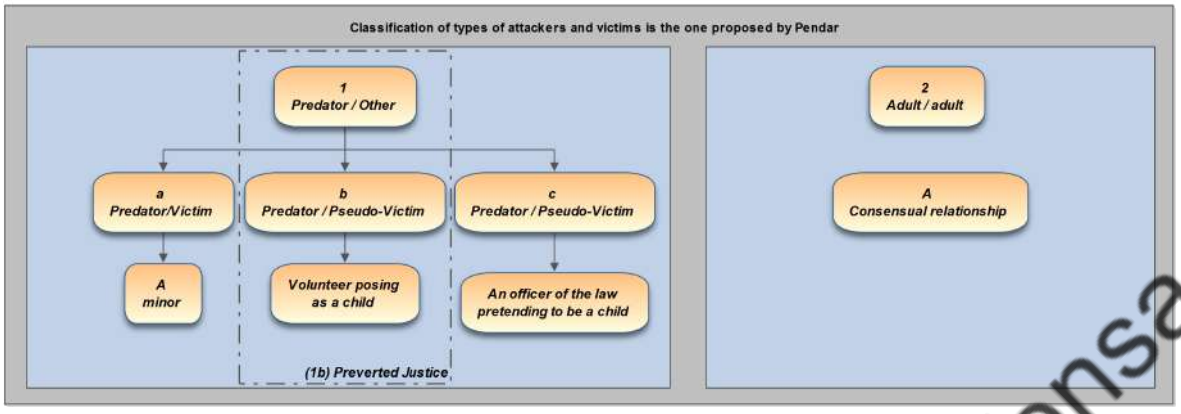


Figure 4.2: Classification of types of attackers and victims [2].

sal faces major problems of data acquisition considering that none of which is necessarily technical. Given these difficulties, the best option is type (1b) interactions that are available online [31, 32].

The website www.perverted-justice.com, which is run by a group of volunteers, aims to make it difficult for pedophiles to take underage victims online. On this website, volunteers are recruited to pretend to be minors (usually from 10 to 15 years old) in chat rooms. When a pedophile has been found, the website publishes online files of all chats with them. In this research, the authors decided to use the aforementioned data to evaluate the feasibility of developing a computer system to perform the automatic recognition of sexual attackers online with the use of type (1b) text records. There, it was evidenced that they managed to distinguish automatically between the pseudo-victim and the attacker, with the assumption that a positive result would support the hypothesis that it is possible to mark suspect chats online automatically [33, 34].

In [29], Bogdanova et al. determined certain distinctive features of pedophiles online, where around 94 % were men who mostly had feelings of inferiority, isolation, loneliness, low self-esteem and emotional immaturity. From this group of criminals, between 60 % and 80 % suffer from other psychiatric illnesses. In general, pedophiles are less emotionally stable people than mentally healthy people. The research referenced by Hall et al. in [3] classifies male pedophiles as

- **homophilia:** if they are only attracted by male children;
- **heterosexual pedophilia:** if they are attracted by girls; and
- **bisexual pedophilia:** if they are attracted by girls and boys.

It also refers to five types of attackers [3]:

- **stalkers:** who approach children in chat rooms to gain physical access to them;
- **cruisers:** who are interested in online sexual abuse and are not willing to meet children off-line;
- **masturbators:** who watch child pornography;
- **networkers or swappers:** who exchange information, pornography and children;

and a combination of the four types. According to their study, the percentage of homosexual pedophiles varies from 9 % to 40 %. The researchers point out that the percentages indicated above do not imply that homosexuals are more prone to attack children, only that a greater percentage of pedophiles are homosexual or bisexual in orientation towards children. As important aspects of this research, the relationship between the age of the victims in relation to the sexual preference of the attacker and the average number of sexual acts is highlighted. Heterosexual male pedophiles prefer children between the ages of 8 and 10 years and on average have performed 34 sexual acts. Homosexual male pedophiles tend to prefer children between the ages of 10 and 13 and on average have performed 52 sexual acts [35, 10]. Regarding bisexual pedophiles, it is only observed that on average they have committed more than 120 acts. In [3], Hall et al. reference a study focused on the incestuous pedophile attacker where the following results were determined: 27 % of all sex offenders assaulted family members. 50 % of crimes committed against children under 6 years were committed by a family member, 42 % of acts committed against children from 6 to 11 years and 24 % against children from 12 to 17 years. An additional study indicates that 68 % of child abusers had sexually abused a family member; 30 % had sexually abused a stepchild or adopted child; 19 % had bothered one or more of their biological children; 18 % had bothered a niece or nephew; and 5 % had sexually abused a grandchild. In this study incestuous heterosexual pedophiles had abused 1.8 children and committed 81.3 acts, while incestuous homosexual pedophiles had abused 1.7 children and committed 62.3 acts.

With regard to the relationship of victims and attackers, Hall's research [3] makes a first distinction:

- **exclusive pedophiles:** pedophiles only attracted to children; and
- **non-exclusive pedophiles:** pedophiles attracted to both adults and children.

In this study, the authors determine that most pedophiles are part of the non-exclusive group.

On the other hand, Vartapetiance et al. in [28] identify three types of potential attackers [36]:

- **strange:** who is a completely unknown person. This type of attacker does not necessarily want to have long-term access to children. Therefore, to attract children, they are more likely to use threats.

- **known:** which can be teachers, drivers, among others. This type of offender generally has access to children; however, they do not use violence to attract them. They invest a lot of time to create the trust relationship to decrease the likelihood of being identified.

- **family:** such as parents, grandparents, cousins, and siblings. This offender generally has long-term access to children, because he is within the family circle and uses his authority to control the children. This type of relationship is the most dangerous due to the time the abuse may last.

In [4], Bogdanova et al. revealed several language characteristics of attackers based on pedophile conversations through chat: implicit / explicit content. On the one hand, the attackers gradually change the context of a conversation until they get an openly sexual conversation without hiding their intentions. First of all, they start with comfortable talks for the victim, which are accompanied by compliments, childish behavior and jargon. Another characteristic evidenced is the fixed discourse, where the attackers use various conversational strategies to avoid departing from the sexual conversation obtained. On some occasions these attackers manipulate their victims by transferring responsibility and blaming them for any differences or disagreements they have had. To minimize the risk of being prosecuted before the law, some attackers force their victims to eliminate chat conversations or records that have been generated. However, it has also been shown that some attackers cease to be cautious and insistently request physical encounters without measuring the consequences [37].

As a summary of the studies related to the understanding of cyber-attackers, in the scope of pedophilia, Figure 4.3 outlines the psychological/technical profile.

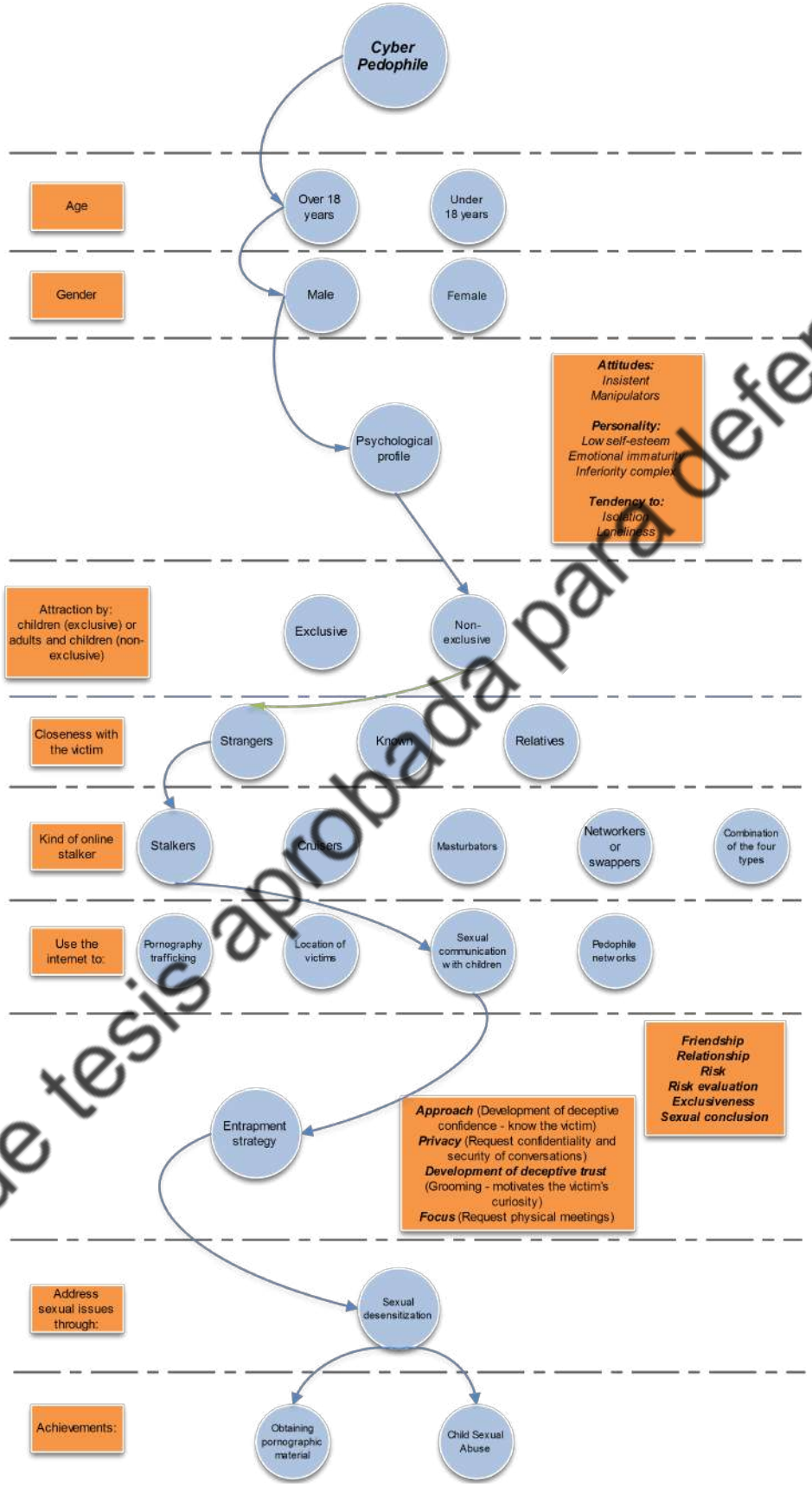


Figure 4.3: Psychological/technical profile of a cyber-pedophile.

4.5 RELATED WORK

There are many works related to online pedophilia and grooming, which concern machine learning paradigms. In 1997, Durkin [5] determined, in one of his researches, that one field of study related to online pedophilia, is the “location of victims through chat rooms.” From this, the author raised several important contributions with the participation of PJ Foundation. The content provided by this agency allowed the application of natural language processing techniques, artificial intelligence and other technological tools associated with machine learning. The results related to the text analysis has allowed scientists to determine certain features of psychological behavior of attackers in relation to the use of technological tools to access their victims [38, 30, 19].

In [27], Bogdanova et al. address the problem of detecting pedophiles with Natural Language Processing techniques (NLP) and the naive bayes and support vector machines classifiers. This problem becomes even more challenging due to the specificity of the chat data. Chat conversations are very different not only from the written text, but also from other types of interactions in social networks, such as blogs and forums, since chat on the Internet usually involves very fast writing. The data usually contains many errors, spelling mistakes, specific jargon, character flooding, among others. The authors also point out the complexity at the time of processing the data with automated parsers. They include a list of features, which includes feelings and other characteristics based on the content. In their experimental results they describe that the classification based on their characteristics can discriminate pedophiles from non-pedophiles with great precision [29, 14, 28].

In a later investigation, Bogdanova et al. [27] propose to model the obsessive discourse of an attacker using lexical chains as a potential feature in the automatic detection of online sexual attackers throughout the conversation. To estimate semantic similarity, they used two parameters: the similarity of Leacock - Chodorow and Resnik. In their results they show considerable variation in the length of lexical chains related to sex according to the nature of the corpus or database. The lexical chains related to sex in the NPS corpus are much shorter, regardless of the similarity of the measure used. The chains in the corpus cybersex are even longer than in the corpus of PJ Foundation. With this premise, they support their hypothesis that this could be a valuable feature in an automated pedophile detection system.

In [14], Cano et al. used a collection of features that aim to characterize attacker conversa-

tions in stages of online preparation through the profile of an attacker based on the characterization of: 1) bag of words (BoW); 2) syntactic; 3) polarity of feeling; 4) content; 5) psycho-linguistic; and 6) speech patterns. The main contributions of this article are: (1) proposal of an approach to automatically identify the stages of preparation in an online conversation based on multiple characteristics: lexicon, syntactic, feeling, content, psycho-linguistics; and patterns of discourse. (2) Classification models for each stage, using unique and multiple characteristics. For the generation of the models, they use several software tools, and the use of the child preparation stages proposed by Olson. In their findings, the authors show that the use of the characteristics of the speech pattern alone can achieve on average a gain on the lexical characteristics. (3) Analysis of particularities to identify the most discriminatory characteristics in each stage of grooming.

In [29], the authors suggest a list of high-level features and study their applicability in the detection of cyber-pedophiles. For this purpose, they used a corpus of downloaded chats from PJ Foundation and two sets of negative data of a different nature: cyberspace records available online and the NPS chat corpus. In their analysis, the authors consider that lexical chains are appropriate for modeling the obsessed speech of pedophile chats. To find semantically related terms, they used parameters of semantic similarity. In particular, the similarity of Leacock and Chodorow and the resemblance of Resnik. The results of the research show that NPS data and pedophile conversations can be accurately discriminated against each other with n-grams (characters), while in the more complicated case of cybersex records high-level characteristics are needed to reach good levels of precision.

Pranoto et al. in [19] try to establish a mathematical logistic model to classify whether an online conversation is a preparation conversation or not. For this purpose, the authors analyzed approximately 160 chat conversations to determine the characteristics of a preparation conversation. These scripts are obtained in a random way from http://www.perverted-justice.com and www.literotika.com. The characteristics are divided into 20 types. The scripts are divided into two sets: 100 scripts for the training set and 59 scripts for the test set. As a result of the research, five most relevant grooming characteristics were identified, and a logistic model was established on this basis. The model is evaluated using the test data set and the results show that the model has acceptable results by the authors.

All the papers present proposals for the analysis of chat conversations with childish pornographic content. It considers different topics that have enable to outline the behavior of an attacker [39]. These topics have addressed aspects of feelings, characteristics based on

content, modeling of obsessive discourse using lexical chains, among others. On the other hand, the stages of grooming have been analyzed in syntactic aspects, polarity of feeling, content, psycho-linguistic and discourse patterns.

4.6 RESEARCH METHODOLOGY

Aiming at determining the life cycle of grooming, we will define stations or phases named *topics*. In the field of text mining is the topic modeling, which allows to analyze a large number of unstructured texts. There are several methods of topic modeling, among the most relevant are Latent Semantic Analysis (LSA), Probabilistic Latent Semantic Analysis (PLSA), Correlated Theme Model (CTM), and LDA. In related literature that compare the performance of LDA with other models in terms of perplexity, it is determined that the performance of LDA is higher than that of other models. Also, it is established that LDA could be applied successfully in various applications aiming at identifying topics in scientific publications, text classification and collaborative filtering [40, 41, 42, 43]. Under these criteria and based on the nature of our study (text categorization), we decided to use LDA as topic modeling.

First, the LDA generative statistical model is proposed; it allows the modeling of topics. Based on this, two experiments were carried out. In the first experiment, several topics are determined according to the characteristics of the pre-processed data. To obtain the data and its processing, the recommendations of the CRISP-DM methodology were followed [18]. After determining an optimal number of topics, we proceeded to give them a logical context through experiment 2. It uses several studies concerning linguistics and communicational intentions to order the topics determined by LDA. Within this ordering, several proposals of life cycles of APT with the topics were related, thus determining the life cycle of the grooming.

4.6.1 Computer attack evaluation

A computer attack represents any hostile activity against a system or a person, using computer applications or psychological persuasion techniques. Every attack has a target, and the responsibility of scientists is to determine what they are in order to apply defense strategies. It is worth noting that computer attackers are aware of the development and execution of each attack by developing a series of phases, stages, or steps to follow to make an attack successful [44, 45, 46, 47].

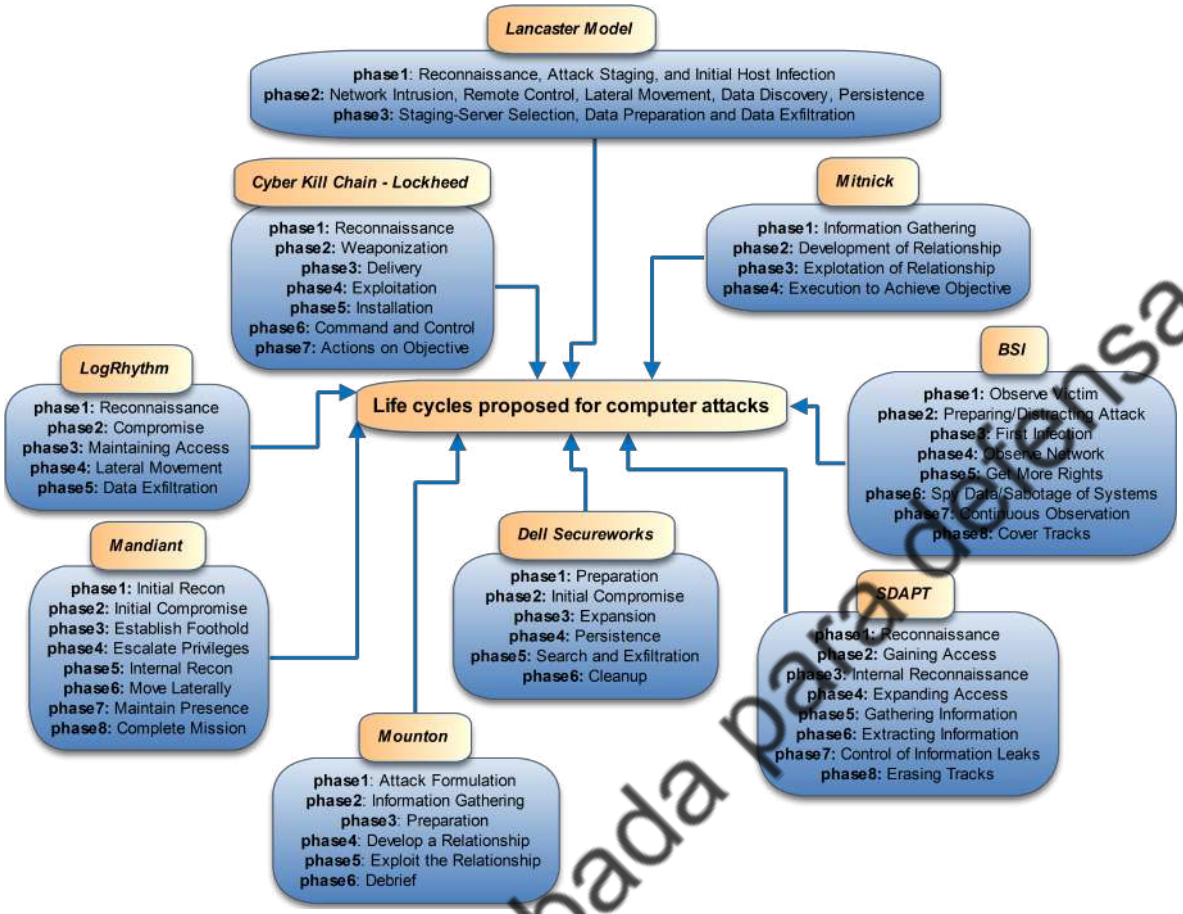


Figure 4.4: Life cycles proposed for computer attacks [47, 52, 53].

To identify the effects of each attack, they should be interpreted not only as isolated incidents or intrusions but also as operations that, in some cases, contemplate long periods. The stages of a computer attack are represented by models of life cycle applicable to cyber-attacks as illustrated in Figure 4.4; they are also known as “cyber-attack chains” [48, 49, 50]. In the scientific field, several authors take as reference the life cycle approach of Lockheed Martin [51], who developed an initial model of cyber-attack chain. Under this criterion, the main contribution of this study is the theoretical/practical definition of the life cycle of grooming, from the point of view of information security.

4.6.2 Research questions

Formalizing the concept of grooming within the field of information security will allow researchers to support future research related to social engineering with the contributions generated with grooming. To achieve these objectives, the following research questions are formulated:

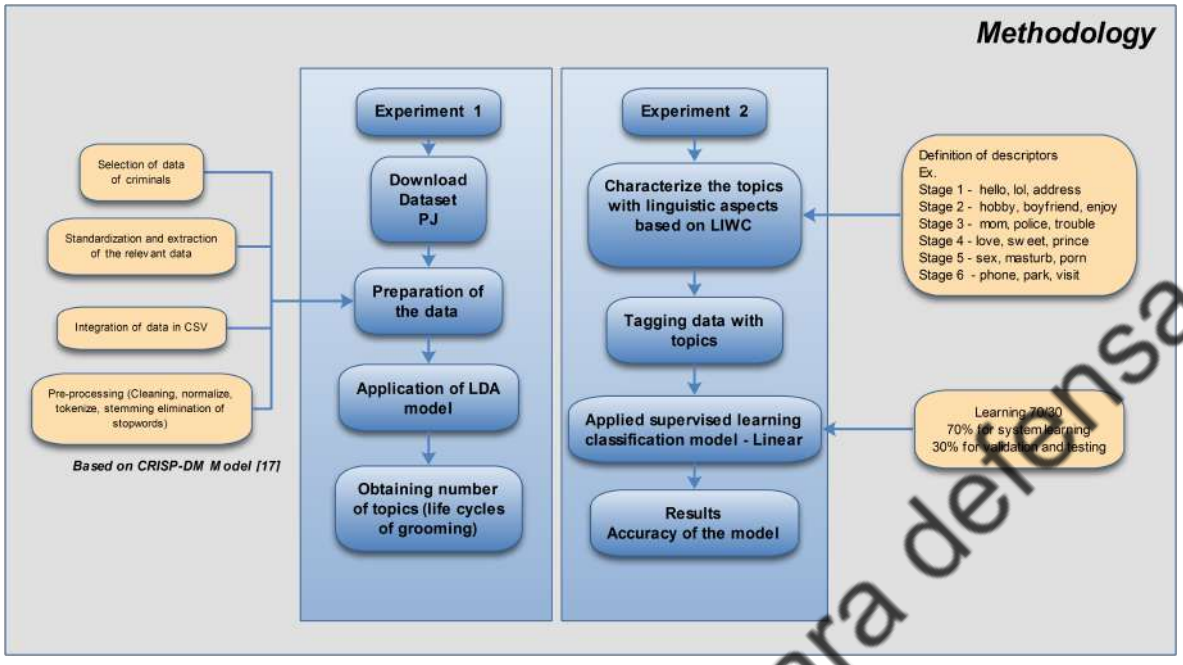


Figure 4.5: Methodology applied to determine the life cycle of grooming.

RQ1: With the use of computer learning, can the phases of grooming be determined as a computer attack?

RQ2: Can grooming be considered an attack vector within the APTs?

RQ3: Can the studies related to grooming support future research associated with social engineering?

4.6.3 Experimental approach

Figure 4.5 describes our experimental approach in two stages. In the first stage, we pre-process the data obtained from PJ, then applied a topic modeling (unsupervised learning paradigm) that allowed us to analyze several phases, stages or steps herein *topics*. Due to the nature of the system development, it should be noted that, in this assignment of topics, linguistic analysis is not previously carried out. After this, we contrasted the results obtained with statistical modeling to justify the exact number of topics that will be applied to grooming. Once this information is obtained, we analyzed if these topics correspond to the different stations of the life cycles proposed in the analysis of computer attacks. In the second stage, we assigned a linguistic approach using a set of word categories provided by the Linguistic Inquiry and Word Count software (LIWC) at each stage determined with LDA. LIWC is a program that analyses text. It reads a given text and counts the percentage

of words that reflect different emotions, thinking styles, social concerns, and even parts of speech. Once the stages were determined, the system was trained with a linear classification model (supervised learning paradigm) to determine the accuracy of the system.

In both experiments, quantitative and qualitative characteristics were adopted (linguistic assignment); however, in the first experiment, the results obtained by the software were justified with statistical analysis. Within the first experiment, the data was pre-processed with the development of scripts with regular expressions to standardize the format of the data. The treatment of the information was applied to 100 chats of pedophile character with an average of 1200 lines of text per chat, processing a total of chat lines of 128171. The number of chats was determined based on the average of conversations analyzed in similar investigations and rejecting short content conversations.

4.7 EXPERIMENTATION

From the scientific method, reproducibility is an essential aspect to be considered. Therefore, each of the phases of the proposed methodology is detailed in the experiments carried out. The data, hardware, and software resources used in the experimentation phase are described in Table 4.2.

4.7.1 Experiment 1

In this section, several aspects that were considered in the realization of the first experiment are addressed. Within these aspects, the obtaining and processing of the data, the application of the LDA model, and the life cycle of grooming are explained.

4.7.1.1 Data collection

100 conversations (128171 chat lines) were downloaded individually in HTML format. The download in this format allowed to use its components (labels) for the pre-processing of the data. The process that was conducted for data collection was:

Table 4.2: Materials Used in Experimentation

Hardware Resources		
Resource	***Description***	***Version***
Processor	Intel Core i5-2320 @3.00 GHz	N/A
Installed memory (RAM)	10,0 GB	N/A
OS	64 bits, x64 processor	N/A
Software Resources		
Resource	***Description***	***Version***
Windows	It is an operating system produced by Microsoft as part of its family of Windows NT operating systems.	N/A
Perl	It is a programming language suitable for writing simple scripts and complex applications	Perl 5 Version 28 Sub version 1 (v5.28.1)
MATLAB	It is a numerical computing environment with multiple paradigms and a programming language developed by MathWorks.	R2019a Version 9.6
Text Analytics Toolbox	It is a MATLAB module that provides tools to perform data mining and machine learning	R2019a Version 9.6
Microsoft Excel	It is a spreadsheet developed by Microsoft that has graphic tools, dynamic tables and a programming language for applications.	Excel 2017 Version 15.0
Data Resources		
Resource	***Description***	***Version***
Perverted Justice	It is the largest virtual repository of grooming conversations publicly available for analysis.	http://www.perverted-justice.com

- **Dataset Download from PJ:** As previously mentioned, 128171 lines of chats downloaded from PJ were used. These records were generated between attackers and pseudo-victims, and their selection was performed based on the representativeness of the data with a manual download for further analysis. The number of records represented the average of data used in related investigations. Conversation records were downloaded individually in HTML format. This format will allow better processing of the data for further analysis.

- **Description of the data:** The superficial properties of the acquired data were examined, and the number of message lines contained in each chat and the number of words contained in each message line were determined.

- **Data exploration:** In this phase, it was identified that all the conversations had a common structure made up of four components: name of the sender (attacker or pseudo-victim), time stamp, message, and annotations of the pseudo-victim. From this analysis, it was determined that the essential components for the proposed study are the name of the sender and the message. The components such as timestamp and annotations were not considered in the study.

- **Verification of data quality:** Determining the quality of the acquired data was challenging because the chats come from various sources, use informal language, the vocabulary consists of slangs, shorthand, emoticons, and contain spelling errors. For this reason, further purification was required, which will be described in the cleaning and pre-processing section.

4.7.1.2 Data processing

Within this stage, exploration and verification of the data to be analyzed was carried out. In a previous analysis, it was possible to identify a typical structure in them, which is composed of four parts: the name of the sender (attacker or pseudo-victim), the time stamp, the message, and the annotation (optional description). It is determined that two of the identified parties are essential for the analysis; these are the name of the sender and the message. The remaining two parts, which are the time stamp and the optional annotations, are not relevant to the present case study. The verification of the quality of the data was a challenging aspect given the nature of the chats, because they come from different sources and the language used in the conversations is extremely informal to contain slangs, shorthand errors, emoticons, and misspellings. Therefore, the data requires several pre-processing and cleaning steps before the analysis can be performed:

- **Data selection - Attackers only:** The decision on what data should be used for the analysis is based on several criteria, including their relevance to the objectives of data mining, as well as technical and quality limitations, limits on the volume of data, and types of data [52]. As described in the previous sections, the data records for this case study are conversations between attackers and pseudo-victims (undercover agents). After analyzing the dialogues, it is evident that the attackers lead the conversations and choose the topic of discussion; most of the time, they force the pseudo-victims to answer unethical questions. The pseudo-victims, in most cases, follow the topic of the conversation with typical answers like “yes”, “no”, “maybe”, “we will see”, among others. Therefore, to effectively analyze the grooming, it was determined that only the messages of the attackers would be analyzed.

- **Standardization and extraction of relevant data:** To extract only the required data, a distinction was generated between the attacker's messages, the pseudo-victim's messages, and time-stamps, making use of the HTML tags. Using perl-based scripts and regular expressions, only the data relevant to the investigation was filtered.
- **Integration of the data in a CSV file:** By having several files of independent conversations, a unified structure is created from the extracted data; this structure includes all the messages coming from the attackers. For doing so, the data is grouped into a file of type CSV, which is constituted as the base structure that contains all the text data to be treated later.
- **Pre-processing:** In the pre-processing stage, two intermediate threads called cleaning and text standardization are performed. For cleaning the text, punctuation marks and special characters and words that add noise to the study are eliminated. These words are known as stop words and are all those articles, prepositions, conjunctions, pronouns, among others, that do not add meaning to the investigation. After cleaning, the standardization stage is performed where all the text is lowercase, the verbs are taken to their base form, for example "getting" to "get," lemmatization and normalization techniques are applied, to finally eliminate all words that have 2 or less characters or that exceed 15 characters.

The obtained results are illustrated in Figure 4.6.

4.7.1.3 Application of LDA model

LSA [53] and LDA [17] are widely used in NLP applications for similar tasks. These methods use semantic distances or similarities/relationships between terms to form cliches or word chains. LSA and LDA use the joint frequency of the concurrency of words in different bodies, and links between them to find closely related words. Although these methods can be used in a similar way for several NLP tasks such as text summary, answer to questions or topic detection, each one uses different measures and has different meanings. LDA generates topical threads under an earlier Dirichlet distribution, while LSA produces a correlation matrix between words and documents. Under this consideration, LDA has been taken as reference for the determination of topics.

Tests of the LDA model: As a first step, the LDA model required a dataset sectioning (90 %

Figure 4.6: Word cloud of pre-processed and processed text.

- 10 %) to evaluate the quality of adjustment (perplexity vs. time) and be able to determine an optimal number of topics. Note that, in the application of the LDA model sectioning was not required, the model was adjusted to the number of defined topics, through the creation of a bag of words with unigrams, obtaining a classification of words by topic as can be seen in Figure 4.7.

4.7.1.4 Obtaining number of topics (life cycle of grooming)

To determine the number of standard topics of grooming, in contrast to the life cycles of computer attacks described in Figure 4.4, we proceeded to choose a range of values that contains several numbers of topics, and in its analysis determine an optimal compromise solution based on the perplexity and processing time in the application of the model. To demonstrate the effects of the compensation, the quality of adjustment, and the adjustment time are calculated. If the optimal number of topics is high, a lower value can be chosen to speed up the adjustment process and determining the most appropriate number of topics allowed. The range considered to determine an optimal compromise solution started with two (topics proposed by Lancaster) and ended with eight topics proposed by Mandiant, BSI

Figure 4.7: Word cloud by topic - LDA model.

and Sdapt [47].

The pre-processed CSV dataset was required to execute the LDA algorithm. The algorithm itself required that the dataset be divided into two groups to train and validate the model. In our case study, 90 % for training and 10 % for validation. The model by its natural defined two groups of bags of words with unigrams.

The optimal value result determined that the number of topics suitable for the analysis of the life cycle of grooming is six, as illustrated in Figure 4.8. As can be seen, the perplexity and the time elapsed for this number of topics is reasonable. Besides, it can be deduced that an increasing number of topics leads to a better adjustment, but adjusting the model takes longer to converge. As additional data, it could be determined that two additional theoretical topics are not testable through the dataset since the first would define the way attackers look for their victim and as a second topic is the demonstration of the mechanisms or associated techniques to maintain contact after performing sexual encounters. This is because the pseudo-victims cease to have communication with pedophiles once they pose fortuitous encounters.

Operation of the LDA model with 5 text lines from the dataset: To analyze the functioning of the LDA model, we proceeded to evaluate it with 5 lines of text from the dataset: 1.-

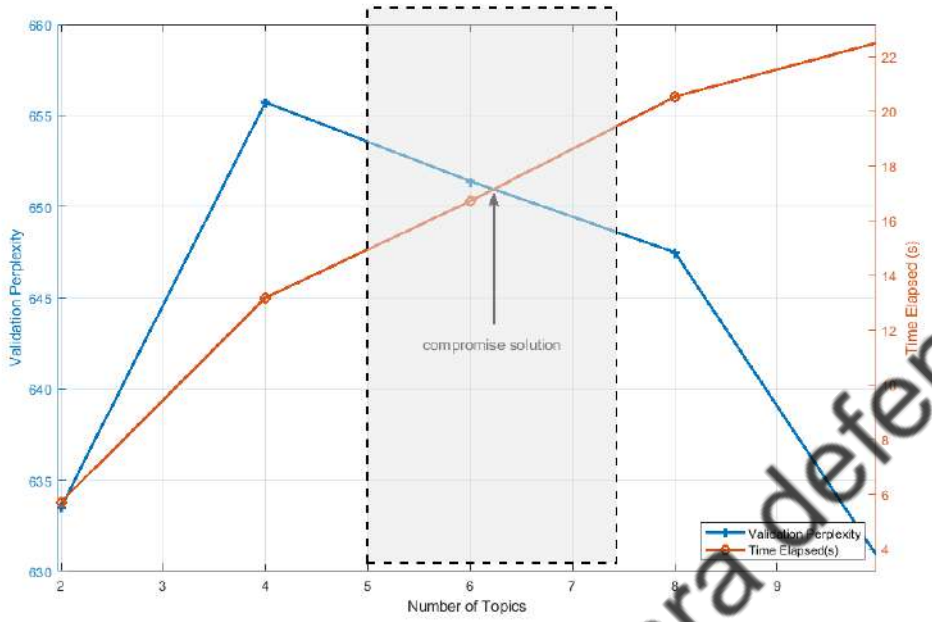


Figure 4.8: Topics assigned to grooming by LDA.

"love give massage", 2.- "nice warm lotion body", 3.- "love", 4.- "give nice ass rub", and 5.- "yeah why". In Figure 4.9, we can observe the mixture of the six topics present in each one of the lines of text analyzed with the LDA model. It is also observed that, for each analyzed document, the probability of a topic stands out above the others, this allows to infer the belonging of the document to a particular topic, understanding a document as a text message and each topic as a phase of grooming.

Operation of the LDA model with test text independent of the dataset: To test the accuracy of the unsupervised LDA model, a separate string of text was created from the data of the dataset ("this will be our little secret ... do not tell your parents about me ... I can get in trouble") obtaining the result depicted in Figure 4.10. In the bar chart, it is observed that in the text used, there are multiple mixtures of topics, and the highest probability of belonging for this text is found in topic number 3.

4.7.2 Experiment 2

In this section, the number of topics determined with linguistic characteristics is related. In this way, a supervised classification model will be applied, and it will be finalized with the analysis of the results and its accuracy.

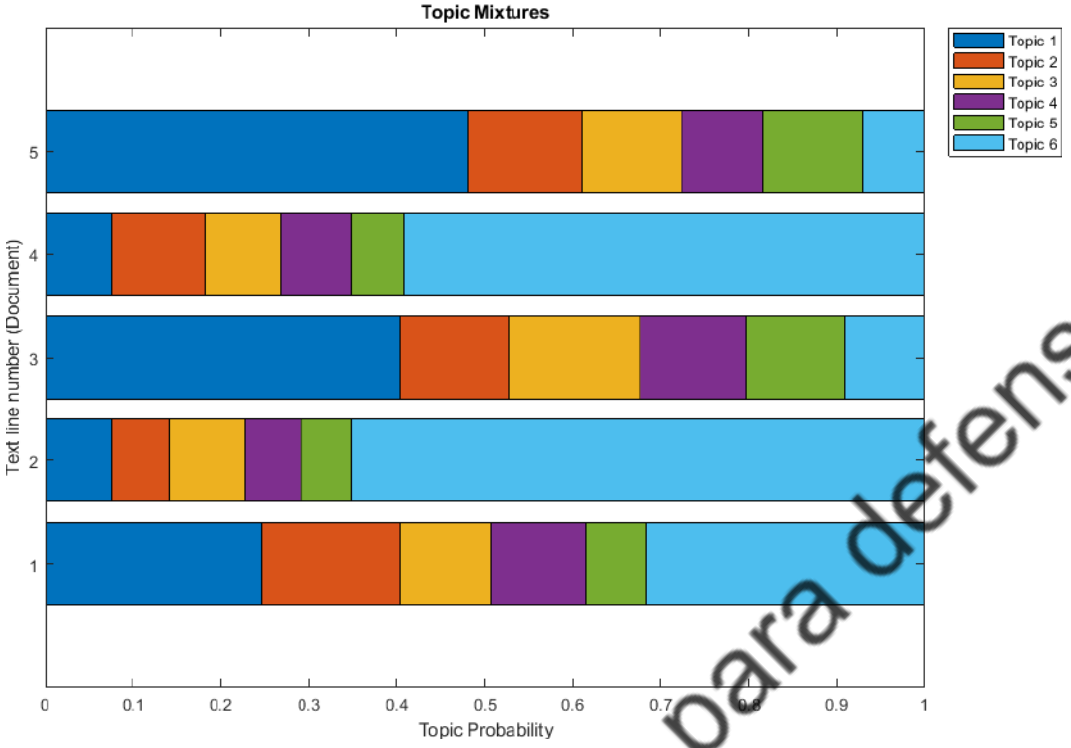


Figure 4.9: Analysis of the operation of the LDA model.

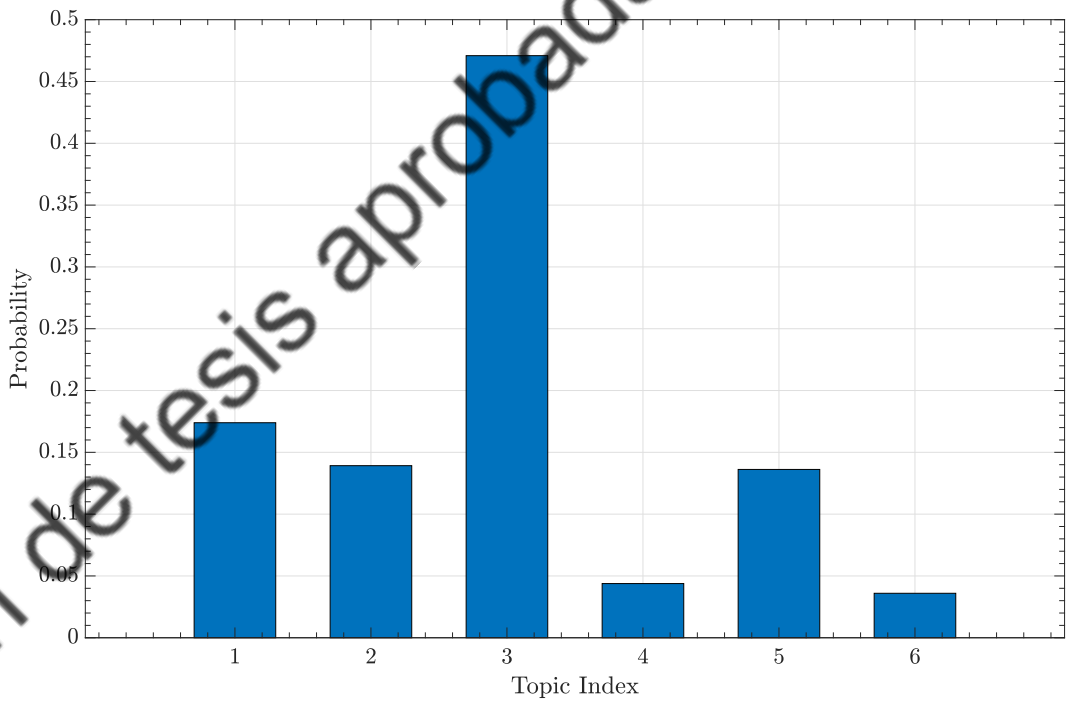


Figure 4.10: Document topic probabilities using the LDA model.

4.7.2.1 Characterizing the topics with linguistic aspects LIWC

As illustrated in Figure 4.11, part of the process that will characterize and define the stages of the life cycle of grooming, as a first phase it is required to combine the topics obtained

with the application of the LDA model, with the categories linguistics provided in [54] and obtained from the LIWC software.

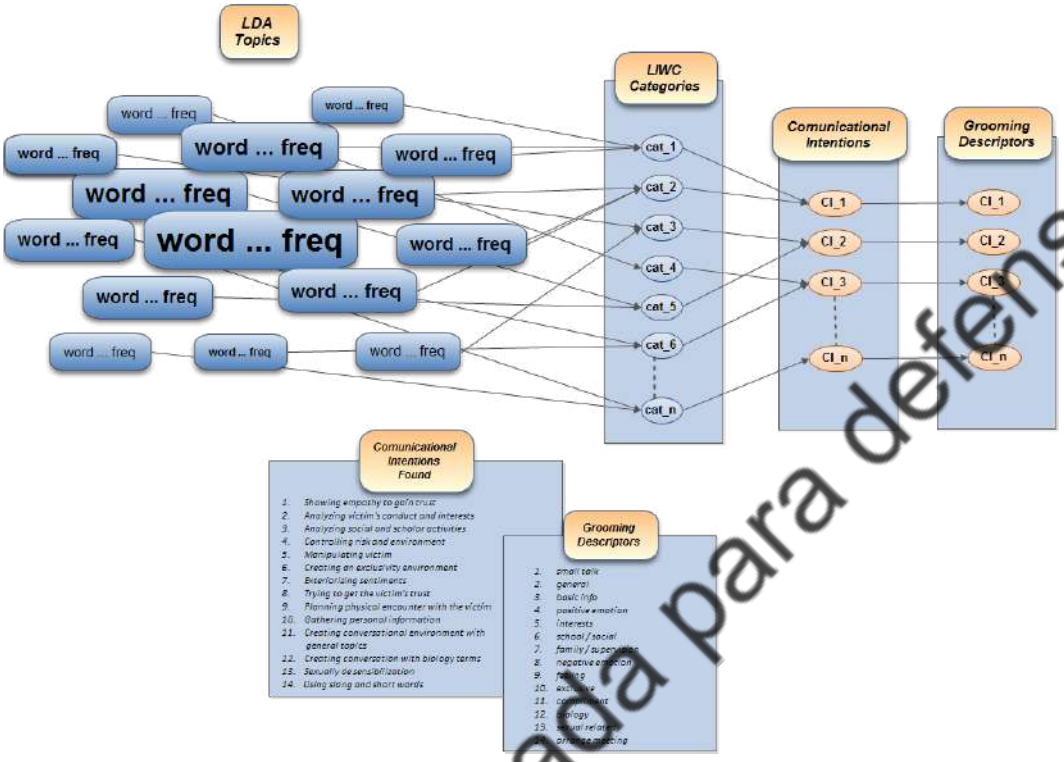


Figure 4.11: Process to characterize and define the stages of the life cycle of grooming.

Proposal of descriptors based on communicational intentions: Applying the LDA model with 6 topics grouped words with their respective frequency within each topic; however, they are only words and require a linguistic process, in order to give a meaning to each topic. Based on Figure 4.11, we will explain the process of characterization of the text of a topic with linguistic aspects. This process was replicated in the rest of the topics as follows

- Obtaining words frequently from each topic;
- The linguistic categories provided by LIWC were compiled as depicted in Figure 4.12;
- Each word was placed within one or more linguistic categories;
- With the linguistic characteristics obtained, it was possible to infer the communicator intentions of the attacker in the selected topic;
- In the process of determining the “communicational intentions” we observed that in some cases these were repeated, due to the nature of the attack. With this background, we created a structure of descriptors that conceptualized each communal intention formulated;

❖ To finish with the formalization stage of the descriptors, we codified them. It is worth noting that these descriptors are specific to the analysis of grooming and that later they will be used to label the data set, a step prior to the application of a supervised learning model.

LIWC Categories	Language examples
You	You
Friend	Friend, boyfriend, girlfriend, lover
Social	Adult, anyone, party, outsider, fight, story, phone, private, public, gossip
Work	Homework, office, school
Leisure	Art, bands, game, hangout, sport, television, movie

Figure 4.12: Linguistic categories LIWC.

4.7.2.2 Tagging data with topics

With the use of the communicational intentions and the descriptors based on the LIWC categories, within our context, we proceeded to assign the corresponding codes to the structure of each line of text. This is made up of one or several descriptors as shown in Figure 4.13. It should be mentioned that for the code to be successful the system, see Figure 4.14, must validate that each word contributes to a general idea of the line of text and belongs to the established categories otherwise the system will exclude from your labeling. Additionally, the process that was accomplished for the definition of the communicational intentions allowed to refine the delimitation and distinction of the different groups of topics that will be categorized in the following section.

Life cycles of a computer attack applied to grooming: With the purpose of analyzing grooming as a computer attack, we proceeded to verify the relationship of the most representative life cycles with the communicational intentions described in the research. In order to proceed with this phase, first operational concepts (definition) were determined that contemplate the communicational intentions and in turn were assigned a topic or station number (see Table 4.3).

Topics (Stages)	Grooming Descriptors		Communicative intention
	Code	Name	
Topic 1	**D11**	Basic info	Acquire basic information of the victim.
	D12	Slang	Emphasize the social and contextual understanding of the victim with the use of adolescent jargon.
	D13	Small talk	Distract the victim by creating conversation about unimportant or uncontroversial issues.
Topic 2	**D21**	Interests	Establish a link with the victim when talking about their personal interests.
	D22	School / social	Acquire specific information of the victim, related to his friends, family, school and social life.
	D23	Positive emotions	Show compassion and understanding to gain the confidence of the victim.
Topic 3	**D31**	Family / supervision	Inquire about the location, the parent's schedule and the victim's supervision.
	D32	Negative emotions	Ensure the silence of the victim by describing the consequences of revealing the nature of their relationship.
Topic 4	**D41**	Exclusivity	Establish an exclusive relationship with the victim.
	D42	Feelings	Express feelings of love, care and confidence.
	D43	Compliments	Adulate the victim to maintain and increase the level of trust.
Topic 5	**D51**	Biology	Desensitize the victim in the sexual theme, using biological terms.
	D52	Sexual related terms	Detail the sexual acts you want to perform with the victim or past sexual experiences.
Topic 6	**D61**	Meeting arrangement	Plan a personal encounter with the victim.

Figure 4.13: Assigning codes to data.

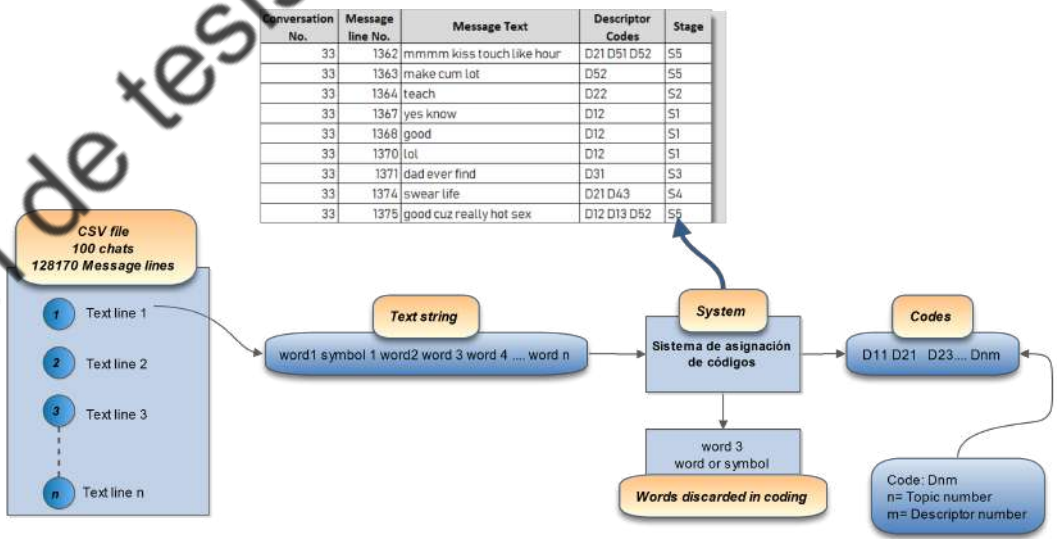


Figure 4.14: Process of data labeling.

Table 4.3: Grooming station

Communicative intention	Grooming Stage	
	Definition	Number
Acquire basic information of the victim.	The attacker makes contact and gets to know his target. He uses short talks to subtly gather general information about the victim, such as his age, gender and interests. This stage can be revisited several times depending on the level of contact the attacker has with the victim.	S1
Emphasize the social and contextual understanding of the victim with the use of adolescent jargon.		
Distract the victim by creating conversation about unimportant or uncontroversial issues.		
Establish a link with the victim when talking about their personal interests.	The attacker establishes a link with the victim by talking about his friends, family, school and social life. The attacker is compassionate and understanding to try to gain the trust of the victim.	S2
Acquire specific information of the victim, related to his friends, family, school and social life.		
Show compassion and understanding to gain the confidence of the victim.		
Inquire about the location, the parent's schedule and the victim's supervision.	The attacker begins to inquire about the location, the parent's schedule and the victim's supervision, using this information to determine the risk of being caught. It encourages the victim not to reveal the nature of their relationship with others and ensures the silence of the victim with various techniques.	S3
Ensure the silence of the victim by describing the consequences of revealing the nature of their relationship.		
Establish an exclusive relationship with the victim.	The attacker tries to establish a trusting and exclusive relationship with the victim. Affirms that they share a special bond. The concept of love, care and feelings in general are introduced.	S4
Express feelings of love, care and confidence.		
Adulate the victim to maintain and increase the level of trust.		
Desensitize the victim in the sexual theme, using biological terms.	When the attacker is sure that the victim trusts him, he becomes more explicit about his intentions. The attacker can ask about past sexual experiences and detail the sexual acts he wishes to perform with the victim. In this stage the sexual content is predominant.	S5
Detail the sexual acts you want to perform with the victim or past sexual experiences.		
Plan a personal encounter with the victim.	It is in this final stage that the attacker attempts a personal encounter with the victim, agreeing on a physical place and date to meet. Additionally, the attacker creates an environment for maintaining the relationship, which allows the attacker to evade any type of detection, safely ending the relationship.	S6

Having determined a characteristic definition of each grooming station, once the communicational intentions have been identified, we proceeded to compare them conceptually with each of the stages of the different life cycles most recognized in the scientific field related to computer attacks within the field of information security as illustrated in Figure 5.13.

- In the analysis of the first grooming station, its correlation was identified with the first station of the life cycle of all the proposals examined. Being **gathering information**, of the SDAP model, which in concept was adapted more to our definition.

- In relation to the second station proposed, the definition found was between stations

Cyber Attack Life Cycles			Grooming Stages					
Author	No	Stages	S1	S2	S3	S4	S5	S6
Lockheed	1	Reconnaissance	x					
	2	Weaponization		x				
	3	Delivery			x			
	4	Exploitation					x	
	5	Installation						
	6	Command & Control (C2)				x		
	7	Actions on Objective						x
LogRhythm	1	Reconnaissance	x					
	2	Compromise					x	x
	3	Maintaining Access						
	4	Lateral Movement			x	x		
	5	Data exfiltration		x				
Mandiant	1	Initial Recon	x					
	2	Initial Compromise		x				
	3	Establish Foothold						
	4	Escalate Privileges					x	
	5	Internal Recon		x				
	6	Move Laterally			x			
	7	Maintain Presence				x		
	8	Complete Mission						x
Dell Secureworks	1	Preparation	x					
	2	Initial Compromise		x				
	3	Expansion				x		
	4	Persistence			x			
	5	Search and Exfiltration					x	
	6	Cleanup			x			
SDAPT	1	Reconnaissance						
	2	Gaining access		x				
	3	Internal Reconnaissance		x				
	4	Expanding Access				x		
	5	Gathering Information	x					
	6	Extracting information					x	
	7	Control of Information leaks			x			
	8	Erasing Tracks			x			
BSI	1	Observe Victim						
	2	Preparing/Distracting Attack		x				
	3	First Infection		x				
	4	Observe Network						
	5	Get More Rights				x		
	6	Spy Data/Sabotage of Systems			x			x
	7	Continuous Observation				x		
	8	Cover Tracks			x			
Mitnick	1	Information Gathering	x					
	2	Development of Relationship		x		x		
	3	Explotation of Relationship					x	
	4	Execution to Achieve Objective						x
Lancaster	1	Reconnaissance, Attack Staging, and Initial Host Infection	x	x				
	2	Network intrution, Remote Control, Lateral Movement, Data Discovery, Persistence			x	x		
	3	Staging-Server Selection, Data Preparation and Data Exfiltration					x	x
Mouton	1	Attack Formulation	x					
	2	Information Gathering	x					
	3	Preparation		x				
	4	Develop a Relationship			x	x		
	5	Exploit the Relationship					x	
	6	Debrief						x
		Grooming Life Cycle	*Gathering Information*	*Gaining Access*	*Lateral Movement*	*Escalating Privileges*	*Execution*	*Debrief*

Figure 4.15: Operational concepts (definition) related to the communicational intentions.

1, 2, 3 and 5 of all the models. However, the second station **gaining access** of the SDAP model defined the station better.

- With regard to the third station, this was identified between stations 3, 4, 6, 7, 8 and 2 of the models analyzed. The fourth station of the Logrythm, **lateral movement**, indicates

the characteristics of the communicational intention that determined this corresponding grooming station.

- ❖ When analyzing the fourth station, the heterogeneity of concepts associated with it was evidenced, being stations 2, 3, 4, 5, 6, 7 of the different authors, those that coincided conceptually with this station. However, the fourth station of Mandiant, **escalate privileges**, was the one that in position and concept defined the station.

- ❖ Stations 4, 2, 5, 6, 3 of the analyzed models showed correspondence with station number 5 of grooming. In this way, Mitnick station 4, **execution to achieve objective**, characterized it more accurately.

- ❖ For the selection of the concept of the last station of the grooming, it was observed that almost in all the stations of the models were contrasted with the last or penultimate station. However, the phase proposed by Mounton et al. [55] is the best one that describes the end of the attack **debrief**, since in this station the attacker manages the mental state of his victim at his convenience with different strategies.

4.7.2.3 Applied supervised learning classification model

For the selection of a supervised learning model classification technique, it is advisable to understand the nature of the problem. This is the case of linear classifiers, given that their simplicity and computational appeal are widely used in problems of automatic text classification, an integral part of our research [56, 57].

Another important aspect of this classification is its usefulness in machine learning and data mining consequently text mining. Unlike nonlinear classifiers, such as neural networks, which allocate data to a higher dimensional space, linear classifiers work directly on the data in the original input space. While linear classifiers cannot handle certain complex data types, they may be enough for textual content data. For example, linear classifiers have been shown to offer competitive returns on document data with non-linear classifiers. An important advantage of linear classification is that the training and testing procedures are much more efficient. Therefore, linear classification can be very useful for some large-scale applications [58, 59, 60].

Below is the process to train a linear classifier that is based on the word frequency count, through a bag-of-words model. Using it as a predictor of the stages of grooming to which a

certain text message belongs.

For the application of the linear model the following steps were followed:

- The pre-processed CSV file is complemented with the labels of the stages of the life cycle of the grooming;
- The dataset is loaded in CSV format to MATLAB;
- A class distribution histogram is constructed to show the presence of each of the grooming stations in the dataset, see Figure 4.16;
- The data set is divided into 2 partitions for training and one set excluded for testing and validation (the training percentage was 70 % and 30 %, respectively);
- The classification model that takes as input a Bag-of-words model, which contains the pre-processed and labeled data, is constructed and trained;
- The classifier is tested to predict the labels of the test data using the trained model and then the classification accuracy is calculated, this being the proportion of labels that the model predicts correctly;
- An array is created with new data (text messages) to test the model.

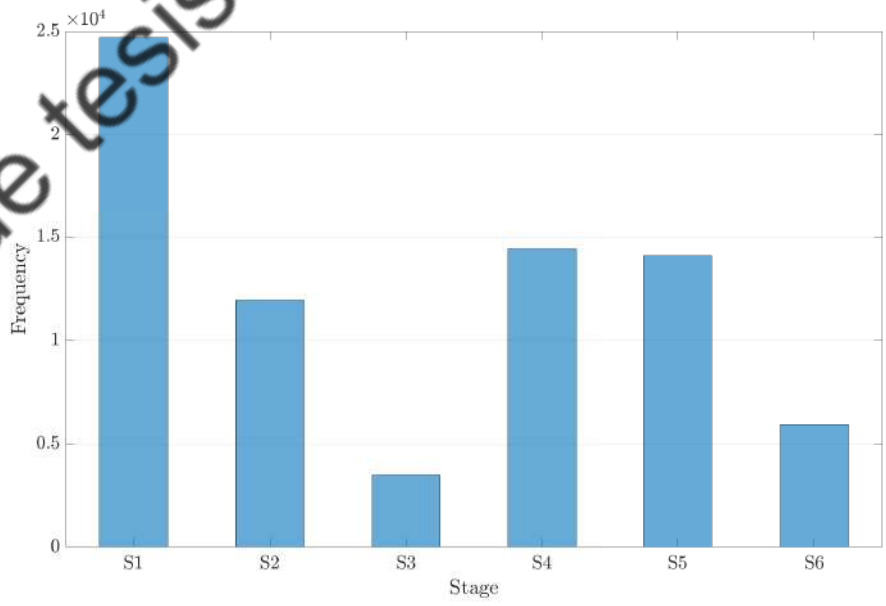


Figure 4.16: Class distribution histogram.

4.7.2.4 Results and model accuracy

After applying the proposed steps sequentially, and applying data that are not known by the system, it describes that its accuracy was 97.61 % (see Figure 4.17), thus confirming that the linear model was adapted without major problems to our case study.

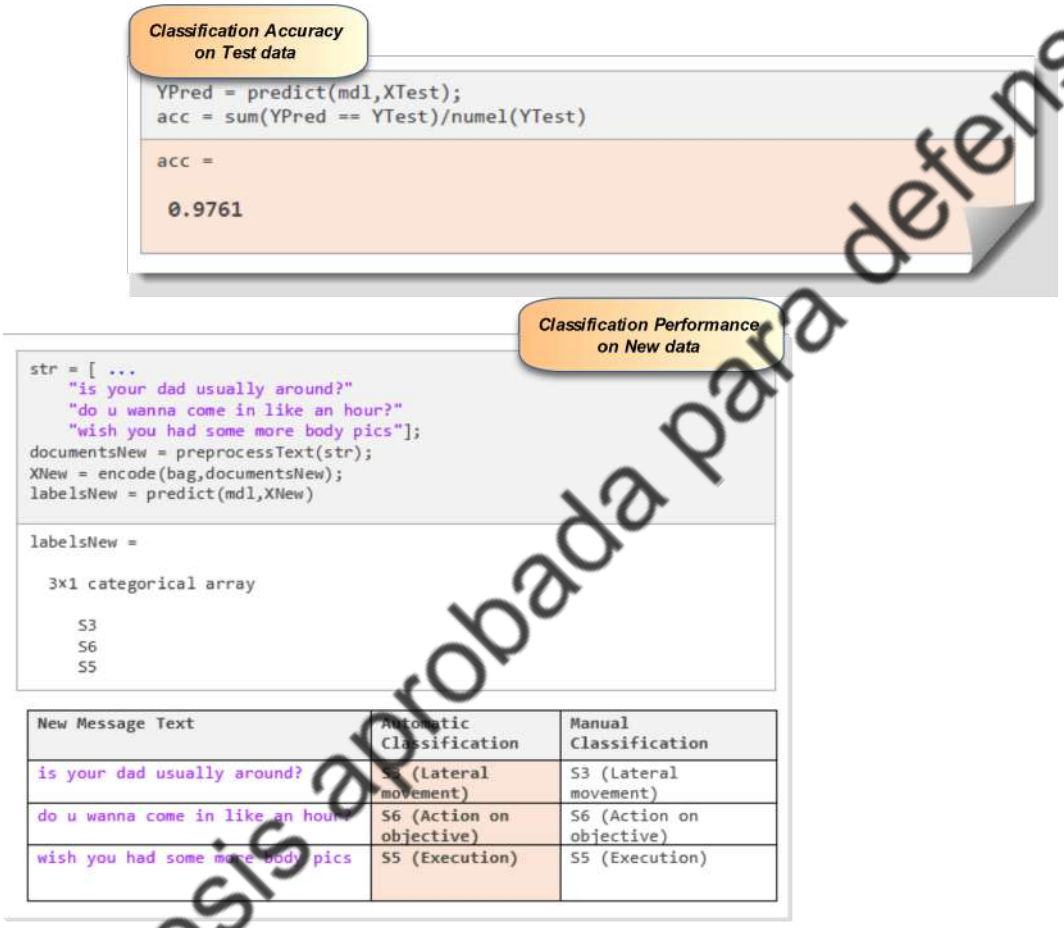


Figure 4.17: Model accuracy and test with new data.

4.7.2.5 Comparison of the proposed linear model with deep learning models

Research related to the detection of cyber-pedophiles and grooming, support their studies with previous investigations related to artificial intelligence. In some cases, they review the literature of implemented algorithms and in other cases they propose new algorithms aimed at improving classification efficiency. Algorithms such as Support Vector Machine (SVM), Naive Bayes, Decision trees and k-nearest neighbor (KNN) and k-means clustering, have already been evaluated [61, 7, 62, 29]. Regarding the deep learning models applied, we tested 2 different models: A convolutional neural network (CNN), see Figure 4.18, and a

Long short-term memory (LSTM) network, see Figure 4.19. The classification accuracies we obtained with these models were 96.11 % for the CNN and 95.91 % for the LSTM network. Based on these results, it is shown that the linear model is the best applied to our case study (classification of texts, see Section 4.7.2.3) since its accuracy is higher (97.61 %) compared with that of the deep learning models.

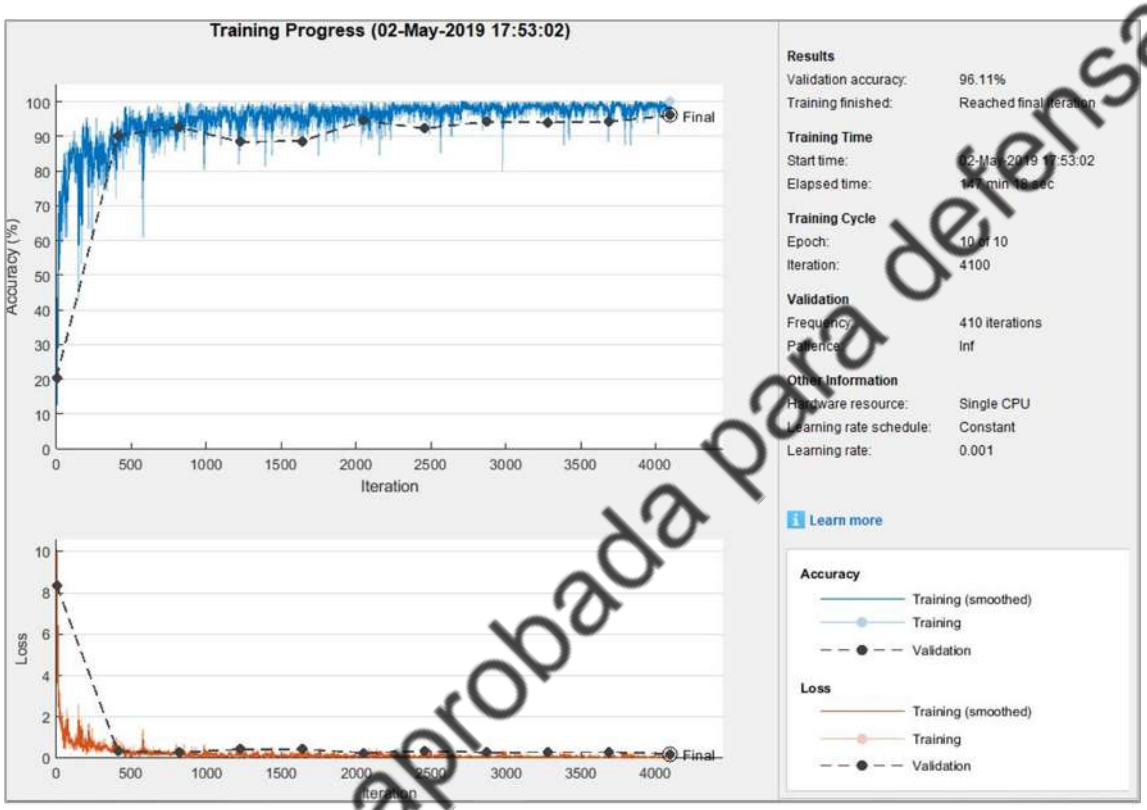


Figure 4.18: CNN network training.

4.8 ANSWERING RESEARCH QUESTIONS

4.8.1 With the use of computer learning, can the phases of grooming be determined as a computer attack?

The researches related to grooming have been analyzed from the psychological point of view, this aspect not being supported by agreements of the scientific community allows researchers to determine different phases with a high degree of subjectivity. For this reason, the research was based on a statistical model LDA that allowed to determine a specific number of stations or phases that attackers follow when applying this attack to their victims. With

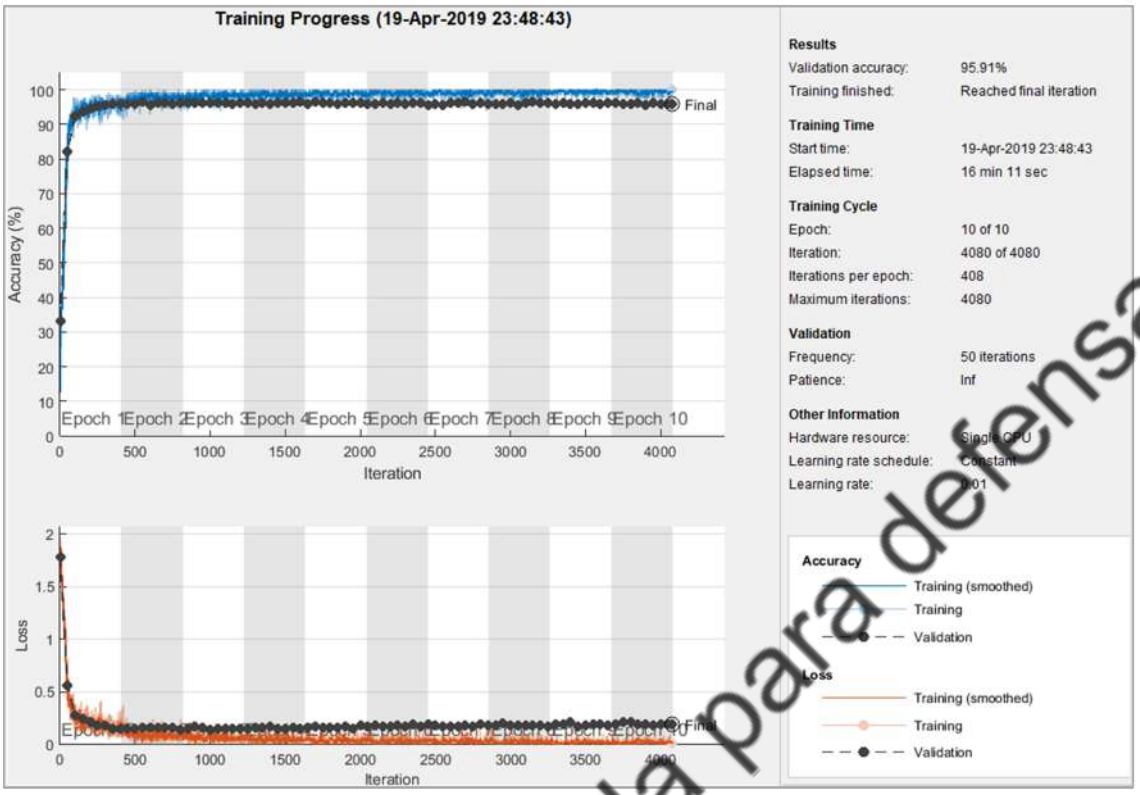


Figure 4.19: LSTM network training.

the application of computational learning it became evident that it is feasible to determine if a text belongs to a specific station with a high degree of accuracy.

4.8.2 Can grooming be considered an attack vector within the APTs?

In the search to be able to place grooming as an attack vector within social engineering and information security, Krombholz et al. [63], in its taxonomy proposal refers to APT. Chen et al. in [46], clearly describes and justifies the difference between APT attacks and traditional cyber-attacks. In this differentiation they determine that the APT come from highly organized, sophisticated, determined and obstinate attackers who direct their attacks to specific people or organizations, government institutions, commercial companies with the purpose of obtaining competitive advantages, strategic benefits that in some situations cause irreparable damage. All this process is successful based on the repetitiveness of their attack attempts, maintaining discretion and non-invasive immediately, but with high resistance capabilities, in the long term in order to meet their objectives.

One of the attacks considered APTs, is social engineering. Attackers who apply social engineering have shown very diverse behavioral patterns (friendship, empathy, threat, abuse of trust, etc.). These psychological traits, when organized in phases, demonstrate a common behavioral pattern that is persistence. As it has been demonstrated in the research, grooming follows this same behavioral pattern, for this reason being a type of social engineering enters the APT classification.

Chowdhary et al. [48] built a list of criteria that determine whether an attack is APT or a common cyber-attack. If the answer to any of these relative criteria is true for the attack case in question, then the attack is not APT. In Table 4.4, a grooming attack is contrasted with the mentioned list of criteria, in this way, it is framed within the group of APT attacks, in order to analyze it from that perspective.

Table 4.4: APT Criteria Applied to Grooming

Criteria	Attack: Grooming	
	Fulfillment	Description
This attack could have been avoided in more than one way	***False***	Given the process of execution, grooming is a surprise, and it is unlikely (concerning the nature of the victim) that it can be avoided with minimal countermeasures and security controls.
This attack did not require much adaptation by the attackers	***False***	The attempts of the attackers, to achieve their objective, require a great adaptation or intense evasive techniques in response to the victim's attempts at defense.
This attack did not show any novelty in its variants.	***False***	Applied social engineering techniques make a grooming attack successful. The novelty in the attack methods used makes it difficult to detect them with existing tools and techniques.

There is a divergence of criteria when updating the concept of an APT; for this reason, it is difficult to take security measures against these unconventional attacks. On the other hand, organizations such as the National Institute of Standards and Technology (NIST) have not taken into account the new objectives and damages caused by the APT. However, the increasing manifestation of the APT with sophisticated methods and deterministic characteristics make the security industry point out the need to review the definition of APT, to include other domains with new attack targets [47].

In the scientific field, several criteria have been proposed to update the concept of an APT. It starts from a military criterion, to refer to a class of sophisticated attacks, carried out by highly skilled attackers, whose objective is to obtain sensitive information from their victim [44]. The definition of an APT is made up of the combination of three terms:

- Threat: the threat in APT attacks is usually the loss of sensitive data, the impediment of critical components or the breaking of the victim. These are growing threats for many national entities and organizations that have advanced protection systems that protect

their data.

- ❖ Persistent: APT attackers are very determined, persistent and obstinate. Once they get access to the victim, they try to extend their stay for as long as possible. They use several evasive techniques to avoid detection and follow a slow and discreet approach to increase their likelihood of success.

- ❖ Advanced: APT attackers usually have advanced tools and methods, necessary to perform an APT attack. These advanced methods include the use of multiple attack vectors to execute, as well as to keep the attack going.

According to the NIST [45], an APT attacker: (i) pursues its objectives repeatedly over a prolonged period of time; (ii) it adapts to the efforts of defenders to resist it; and (iii) is determined to maintain the level of interaction necessary to achieve its objectives. These objectives are usually the theft of information or the deterioration of critical aspects of a mission or program through multiple attack vectors.

Within the study of the APT [47, 48, 49, 51], we have observed the interest of contributing to the detailed study of differentiated attacks according to their processes by applying life cycles proposed by industry and academia (see Figure 5.2) based on concepts of computer attacks.

4.8.3 Can the studies related to grooming support future research associated with social engineering?

Taking into account that the studies of social engineering are in continuous development and have as one of their objectives to determine behavioral patterns of the attackers and their victims, it is evident that the present study can support future investigations aligned to the study of social engineering as an APT within the field of information security.

4.9 CONCLUSIONS AND FUTURE WORK

We have positioned grooming as an attack vector within social engineering and information security. Through the modeling of topics, different stages or seasons of a life cycle of

grooming associated with social engineering is determined; this will allow supporting investigations related to identifying patterns of malicious behavior online. Additionally, a psychological and technical profile of the type of attacker associated with online pedophilia has been presented. We have conducted two experiments, the first consists of determining, in a computational way through a statistical model, stations, or cycles related to grooming. The second experiment gives a linguistic concept to the established stations. In the last experiment, a linear model of machine learning was applied, according to the determined linguistic characteristics, aiming at characterizing text pertinent to the case study (online pedophilia), obtaining an accuracy percentage of 97 %. All data was selected and downloaded from the Perverted-Justice website. It is worth noting that although the related work does not align directly with information security, we address the topic following an information security approach. For this reason, the research covers several fields aligned to security, such as APT persistent advanced attacks and social engineering. The processing and evaluation of short text lines obtained from instant messaging protocol, through the proposed approach, does not only apply to the case study but can be reproduced in other security-related fields, these can be online bullying, bank fraud, phishing, among others. One of the main challenges in the path of new cases of study is obtaining relevant data related to the research field; for this reason, it is essential to promulgate and to disseminate in the scientific community this type of studies, to gain more interest in the academy and industry. As future work, we have planned to implement the model in parental control systems for further optimization. As a step before this implementation, the model must be contrasted with data from instant messaging, with texts of adult conversations of a sexual nature and frequent conversations. In this way, the system will have the ability to unlink these conversations from the classification and location of relevant texts to the violation of privacy.

REFERENCES

[1] Lyta Penna, Andrew Clark, and George Mohay. Challenges of automating the detection of paedophile activity on the internet. *Proceedings - First International Workshop on Systematic Approaches to Digital Forensic Engineering*, 2005:206–220, 2005.

[2] Nick Pendar. Toward spotting the pedophile telling victim from predator in text chats. *ICSC 2007 International Conference on Semantic Computing*, (c):235–241, 2007.

[3] Ryan C.W. Hall and Richard C.W. Hall. A profile of pedophilia: Definition, characteristics of offenders, recidivism, treatment outcomes, and forensic issues. *Mayo Clinic Proceedings*, 82(4):457–471, 2007.

[4] Dasha Bogdanova, Saint Petersburg, Paolo Rosso, and Thamar Solorio. On the Impact of Sentiment and Emotion Based Features in Detecting Online Sexual Predators. *Proceedings of the 3rd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, (July):110–118, 2012.

[5] Keith F Durkin. Misuse of the internet by pedophiles: Implications for law enforcement and probation practice. *Fed. Probation*, 61:14, 1997.

[6] Hugo Jair Escalante, Esaú Villatoro-Tello, Sara E Garza, A Pastor López-Monroy, Manuel Montes-y Gómez, and Luis Villaseñor-Pineda. Early detection of deception and aggressiveness using profile-based representations. *Expert Systems with Applications*, 89:99–111, 2017.

[7] India McGhee, Jennifer Bayzick, April Kontostathis, Lynne Edwards, Alexandra McBride, and Emma Jakubowski. Learning to identify internet sexual predation. *International Journal of Electronic Commerce*, 15(3):103–122, 2011.

[8] Ryan CW Hall and Richard CW Hall. A profile of pedophilia: definition, characteristics of offenders, recidivism, treatment outcomes, and forensic issues. *Focus*, 7(4):522–537, 2009.

[9] Dasha Bogdanova, Paolo Rosso, and Thamar Solorio. On the impact of sentiment and emotion based features in detecting online sexual predators. In *Proceedings of the 3rd workshop in computational approaches to subjectivity and sentiment analysis*, pages 110–118. Association for Computational Linguistics, 2012.

[10] Carmit Katz. Internet-related child sexual abuse: What children tell us in their testimonies. *Children and Youth Services Review*, 35(9):1536–1542, 2013.

[11] Joanna Cahall Young and Cathy Spatz Widom. Long-term effects of child abuse and neglect on emotion processing in adulthood. *Child abuse & neglect*, 38(8):1369–1381, 2014.

[12] Katrinna MacFarlane and Violeta Holmes. Agent-mediated information exchange: Child safety online. In *2009 International Conference on Management and Service Science*, pages 1–5. IEEE, 2009.

[13] Hady Pranoto, Fergyanto E Gunawan, and Benfano Soewito. Logistic models for classifying online grooming conversation. *Procedia Computer Science*, 59:357–365, 2015.

[14] A.E. Cano, M. Fernandez, and H. Alani. Detecting child grooming behaviour patterns on social media. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8851:412–427, 2014.

[15] India McGhee, Jennifer Bayzick, April Kontostathis, Lynne Edwards, Alexandra McBride, and Emma Jakubowski. Learning to Identify Internet Sexual Predation. *International Journal of Electronic Commerce*, 15(3):103–122, 2011.

[16] Ryan Heartfield, George Loukas, and Diane Gan. You are probably not the weakest link: Towards practical prediction of susceptibility to semantic social engineering attacks. *IEEE Access*, 4:6910–6928, 2016.

[17] DM Blei, A Ng, and M Jordan. Latent dirichlet allocation journal of machine learning research (3). 2003.

[18] Steffen Huber, Hajo Wiemer, Dorothea Schneider, and Steffen Ihlenfeldt. Dmme: Data mining methodology for engineering applications–a holistic extension to the crisp-dm model. *Procedia CIRP*, 79:403–408, 2019.

[19] Hady Pranoto, Fergyanto E. Gunawan, and Benfano Soewito. Logistic Models for Classifying Online Grooming Conversation. *Procedia Computer Science*, 59(Iccsci):357–365, 2015.

[20] Siew Yong, Dale Lindskog, Ron Ruhl, and Pavol Zavarsky. Risk mitigation strategies for mobile Wi-Fi robot toys from online pedophiles. *Proceedings - 2011 IEEE International Conference on Privacy, Security, Risk and Trust and IEEE International Conference on Social Computing, PASSAT/SocialCom 2011*, pages 1220–1223, 2011.

[21] Paul Elzinga, Karl Erich Wolff, and Jonas Poelmans. Analyzing chat conversations of pedophiles with temporal relational semantic systems. *Proceedings - 2012 European Intelligence and Security Informatics Conference, EISIC 2012*, pages 242–249, 2012.

[22] Katrinna MacFarlane and Violeta Holmes. Agent-mediated information exchange: Child safety online. *Proceedings - International Conference on Management and Service Science, MASS 2009*, 2009.

[23] Adams Wai Kin Kong. Tutorial-1: New criminal and victim identification methods for sexual offenses against women and children. In *Electrical and Computer Engineering (WIECON-ECE), 2015 IEEE International WIE Conference on*, pages 1–4. IEEE, 2015.

[24] Moshe Rutgaizer, Yuval Shavitt, Omer Vertman, and Noa Zilberman. Detecting pedophile activity in BitTorrent networks. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7192 LNCS:106–115, 2012.

[25] Loreen N Olson, Joy L Daggs, Barbara L Ellevold, and Teddy KK Rogers. Entrapping the innocent: Toward a theory of child sexual predators' luring communication. *Communication Theory*, 17(3):231–251, 2007.

[26] Alfred Hofmann, Ursula Barth, Ingrid Haas, and Frank Holzwarth. Detection of Child Sexual Abuse Media : Classification of the Associated Filenames. pages 1–5, 2013.

[27] Dasha Bogdanova, Paolo Rosso, and Thamar Solorio. Modelling Fixated Discourse in Chats with Cyberpedophiles. *Proceedings of the Workshop on Computational Approaches to Deception Detection*, pages 86–90, 2012.

[28] Anna Vartapetiance and Lee Gillam. "Our Little Secret": pinpointing potential predators. *Security Informatics*, 3(1):1–19, 2014.

[29] Dasha Bogdanova, Paolo Rosso, and Thamar Solorio. Exploring high-level features for detecting cyberpedophilia. *Computer Speech and Language*, 28(1):108–120, 2014.

[30] Fergyanto E. Gunawan, Livia Ashianti, Sevenpri Candra, and Benfano Soewito. Detecting online child grooming conversation. *Proceedings - 11th 2016 International Conference on Knowledge, Information and Creativity Support Systems, KICSS 2016*, 2017.

[31] Benoit Leclerc, Richard Wortley, and Stephen Smallbone. Getting into the script of adult child sex offenders and mapping out situational prevention measures. *Journal of research in crime and delinquency*, 48(2):209–237, 2011.

[32] Catherine D Marcum. Interpreting the intentions of internet predators: An examination of online predatory behavior. *Journal of Child Sexual Abuse*, 16(4):99–114, 2007.

[33] Kimberly J Mitchell, David Finkelhor, Lisa M Jones, and Janis Wolak. Growth and change in undercover online child exploitation investigations, 2000–2006. *Policing & Society*, 20(4):416–431, 2010.

[34] Kimberly J Mitchell, David Finkelhor, Lisa M Jones, and Janis Wolak. Use of social networking sites in online sex crimes against minors: an examination of national incidence and means of utilization. *Journal of Adolescent Health*, 47(2):183–190, 2010.

[35] L Crystal Jiang and Jeffrey T Hancock. Absence makes the communication grow fonder: Geographic separation, interpersonal media, and intimacy in dating relationships. *Journal of Communication*, 63(3):556–577, 2013.

[36] Helen Whittle, Catherine Hamilton-Giachritsis, Anthony Beech, and Guy Collings. A review of online grooming: Characteristics and concerns. *Aggression and violent behavior*, 18(1):62–70, 2013.

[37] Kimberly J Mitchell, Lisa M Jones, David Finkelhor, and Janis Wolak. Understanding the decline in unwanted online sexual solicitations for us youth 2000–2010: Findings from three youth internet safety surveys. *Child abuse & neglect*, 37(12):1225–1236, 2013.

[38] Lyta Penna, Andrew Clark, and George Mohay. A framework for improved adolescent and child safety in MMOs. *Proceedings - 2010 International Conference on Advances in Social Network Analysis and Mining, ASONAM 2010*, pages 33–40, 2010.

[39] Md Waliur Rahman Miah, John Yearwood, and Sid Kulkarni. Detection of child exploiting chats from a mixed chat dataset as a text classification task. In *Proceedings of the Australasian Language Technology Association Workshop 2011*, pages 157–165, 2011.

[40] Wei Wang, Payam Mamaani Barnaghi, and Andrzej Bargiela. Probabilistic topic models for learning terminological ontologies. *IEEE Transactions on Knowledge and Data Engineering*, 22(7):1028–1040, 2009.

[41] Ali Daud, Juanzi Li, Lizhu Zhou, and Faqir Muhammad. Knowledge discovery through directed probabilistic topic models: a survey. *Frontiers of computer science in China*, 4(2):280–301, 2010.

[42] Yue Lu, Qiaozhu Mei, and ChengXiang Zhai. Investigating task performance of probabilistic topic models: an empirical study of plsa and lda. *Information Retrieval*, 14(2):178–203, 2011.

[43] Rubayyi Alghamdi and Khalid Alfalqi. A survey of topic modeling in text mining. *Int. J. Adv. Comput. Sci. Appl.(IJACSA)*, 6(1), 2015.

[44] Shirley Radack. Managing information security risk: Organization, mission and information system view. Technical report, National Institute of Standards and Technology, 2011.

[45] Richard Kissel. *Glossary of key information security terms*. Diane Publishing, 2011.

[46] Ping Chen, Lieven Desmet, and Christophe Huygens. A study on advanced persistent threats. In *IFIP International Conference on Communications and Multimedia Security*, pages 63–72. Springer, 2014.

[47] Brahim ID Messaoud, Karim Guennoun, Mohamed Wahbi, and Mohamed Sadik. Advanced persistent threat: New analysis driven by life cycle phases and their challenges. In *2016 International Conference on Advanced Communication Systems and Information Security (ACOSIS)*, pages 1–6. IEEE, 2016.

[48] Adel Alshamrani, Sowmya Myneni, Ankur Chowdhary, and Dijiang Huang. A survey on advanced persistent threats: Techniques, solutions, challenges, and research opportunities. *IEEE Communications Surveys & Tutorials*, 2019.

[49] Lin Shenwen, Li Yingbo, and Du Xiongjie. Study and research of apt detection technology based on big data processing architecture. In *2015 IEEE 5th International Conference on Electronics Information and Emergency Communication*, pages 313–316. IEEE, 2015.

[50] Martin Ussath, David Jaeger, Feng Cheng, and Christoph Meinel. Advanced persistent threats: Behind the scenes. In *2016 Annual Conference on Information Science and Systems (CISS)*, pages 181–186. IEEE, 2016.

[51] Eric M Hutchins, Michael J Cloppert, and Rohan M Amin. Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains. *Leading Issues in Information Warfare & Security Research*, 1(1):80, 2011.

[52] Gonzalo Mariscal, Oscar Marban, and Covadonga Fernandez. A survey of data mining and knowledge discovery process models and methodologies. *The Knowledge Engineering Review*, 25(2):137–166, 2010.

[53] Thomas K Landauer and Susan T Dumais. A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review*, 104(2):211, 1997.

[54] Pamela J Black, Melissa Wollis, Michael Woodworth, and Jeffrey T Hancock. A linguistic analysis of grooming strategies of online child sex offenders: Implications for our understanding of predatory sexual behavior in an increasingly computer-mediated world. *Child Abuse & Neglect*, 44:140–149, 2015.

[55] Francois Mouton, Mercia M Malan, Louise Leenen, and Hein S Venter. Social engineering attack framework. In *Information Security for South Africa (ISSA), 2014*, pages 1–9. IEEE, 2014.

[56] Aarti Sharma, Rahul Sharma, Vivek Kr Sharma, and Vishal Shrivatava. Application of data mining–a survey paper. *International Journal of Computer Science and Information Technologies*, 5(2):2023–2025, 2014.

[57] Pradnya P Sondwale. Overview of predictive and descriptive data mining techniques. *International Journal of Advanced Research in Computer Science and Software Engineering*, 5(4):262–265, 2015.

[58] Nikita Jain and Vishal Srivastava. Data mining techniques: a survey paper. *IJRET: International Journal of Research in Engineering and Technology*, 2(11):2319–1163, 2013.

[59] Petar Ristoski and Heiko Paulheim. Semantic web in data mining and knowledge discovery: A comprehensive survey. *Web semantics: science, services and agents on the World Wide Web*, 36:1–22, 2016.

[60] Brijesh Kumar Baradwaj and Saurabh Pal. Mining educational data to analyze students'performance. *arXiv preprint arXiv:1201.3417*, 2012.

[61] Amparo Elizabeth Cano, Miriam Fernandez, and Harith Alani. Detecting child grooming behaviour patterns on social media. In *International conference on social informatics*, pages 412–427. Springer, 2014.

[62] Fergyanto E Gunawan, Livia Ashianti, Sevenpri Candra, and Benfano Soewito. Detecting online child grooming conversation. In *2016 11th International Conference on Knowledge, Information and Creativity Support Systems (KICSS)*, pages 1–6. IEEE, 2016.

[63] Katharina Krombholz, Heidelinde Hobel, Markus Huber, and Edgar Weippl. Advanced social engineering attacks. *Journal of Information Security and Applications*, 22:113–122, 2015.

5 UNDERSTANDING CYBERBULLYING AS AN INFORMATION SECURITY ATTACK - LIFE CYCLE MODELING

Patricio Zambrano[1*], Jenny Torres[1], Ángel Yánez[1], Alexandra Macas[1], Luis Tello-Oquendo[2]

[1]Department of Informatics and Computer Science, Escuela Politécnica Nacional, Quito 170525, Ecuador

[2]College of Engineering, Universidad Nacional de Chimborazo, Riobamba 060108, Ecuador

5.1 ABSTRACT

Nowadays, cyberbullying cases are more common due to free access to technological resources. Studies related to this phenomenon from the fields of computer science and computer security are still limited. Several factors such as the access to specific databases on cyberbullying, the unification of scientific criteria that assess the nature of the problem, or the absence of real proposals that prevent and mitigate this problem could motivate the lack of interest by researchers in the field of information security to generate significant contributions. This research proposes a cyberbullying life cycle model through topic modeling and conceptualizes the different stages of the attack considering criteria associated with computer attacks. This proposal is supported by a review of the specific literature and knowledge bases gained from experiences of victims of online harassment and tweets from attackers.

KEY WORDS: Cyberbullying, Pattern Behavior, APT, Social Engineering.

5.2 INTRODUCTION

Cyberbullying is a systemic process where attackers use technology as a communication medium to threaten, embarrass, intimidate, or criticize their victims. Aspects such as threats, texts with rude, aggressive or derogatory connotation and comments published on any Internet media are prevalent in this attack. In some cases, the attackers upload photographs or videos that increase discomfort and shame in their victims [1].

When carrying out cyberbullying, the attackers focus their malicious comments on aspects such as gender, religion, sexual orientation, race, or physical differences of their victims. This attack affects their victims and their families psychologically due to the level of aggressiveness it presents. In some cases, this attack is anonymous and difficult to identify, as well as to control.

Scientific contributions in this field are still limited. The primary motivation of this research is to propose a model applicable to harassment in line with information security concepts, justifying its relevance and being part of future research related to this phenomenon.

For conducting this study, we follow the CRISP-DM methodology [2], which consists of several phases. One of these allowed to know in depth the problem of bullying, the technical and scientific contributions proposed in the literature, potential causes and effects that make the problem a viable case study. In the data recognition, the lack of data analysis provided by the victims and the relationship with the social engineering attacks was evidenced, which is why we proceeded to search for related information to contrast it with information from the attackers (Twitter). Once the databases were processed, cyberbullying was modeled in topics using the Latent Dirichlet Allocation (LDA) model. Once obtained the word groupings that were related to each topic, the linguistic meaning was given to determine the communicational intentions intrinsic to the data. This treatment revealed patterns similar to those already proposed in the studies of Advanced Persistent Threat (APT); therefore, depending on the pattern, each phase was named, thus generating a viable model adapted to information security concepts. In addition to the above, the research analyzes the behavior of two machine learning models, which offered results that enable the development of future preventive defense applications.

The main contributions of this study are summarized as follows.

- ❖ Position cyberbullying as an attack associated with social engineering;
- ❖ Define a process for selecting the data source with content associated with cyberbullying;
- ❖ Propose a model applicable to cyberbullying that uses concepts associated with information security to prevent and mitigate this attack. The model is based on the knowledge bases of attackers and victims.

The remainder of the article develops each phase of the CRISP-DM methodology. Section 6.3 presents the methodology and its phases. Section 6.4 describes the problem based on scientific literature, evaluates the problem from the social engineering perspective, and poses the research questions. Section 5.5 describes the knowledge bases that will be processed. Section 6.5 develops a procedure for data processing. Section 6.6 describes the cyberbullying modeling process. Section 5.8 evaluates the results obtained and implements two machine learning models to analyze the behavior of the model and the data. Finally, Section 5.9 draws conclusions and presents future works.

5.3 METHODOLOGY

This study is based on the guidelines proposed by the CRISP-DM methodology [2], as illustrated in Figure 5.1. This methodology normalizes the knowledge discovery process, data science, or data analytics when large amounts of digital information are analyzed. CRISP-DM recommends five main phases for the treatment of information, namely business understanding (herein problem understanding), data understanding, data preparation, modeling, and evaluation. In what follows, these phases are elaborated, aiming at positioning cyberbullying as an attack associated with social engineering and information security. Through the modeling of topics, different stages or seasons of a life cycle of cyberbullying associated with social engineering is determined; this will allow supporting investigations related to identifying patterns of malicious behavior online.

5.4 PROBLEM UNDERSTANDING

At this phase, a review of the literature is conducted to define the problem and the main scientific contributions. On the other hand, the position of cyberbullying within the field of in-

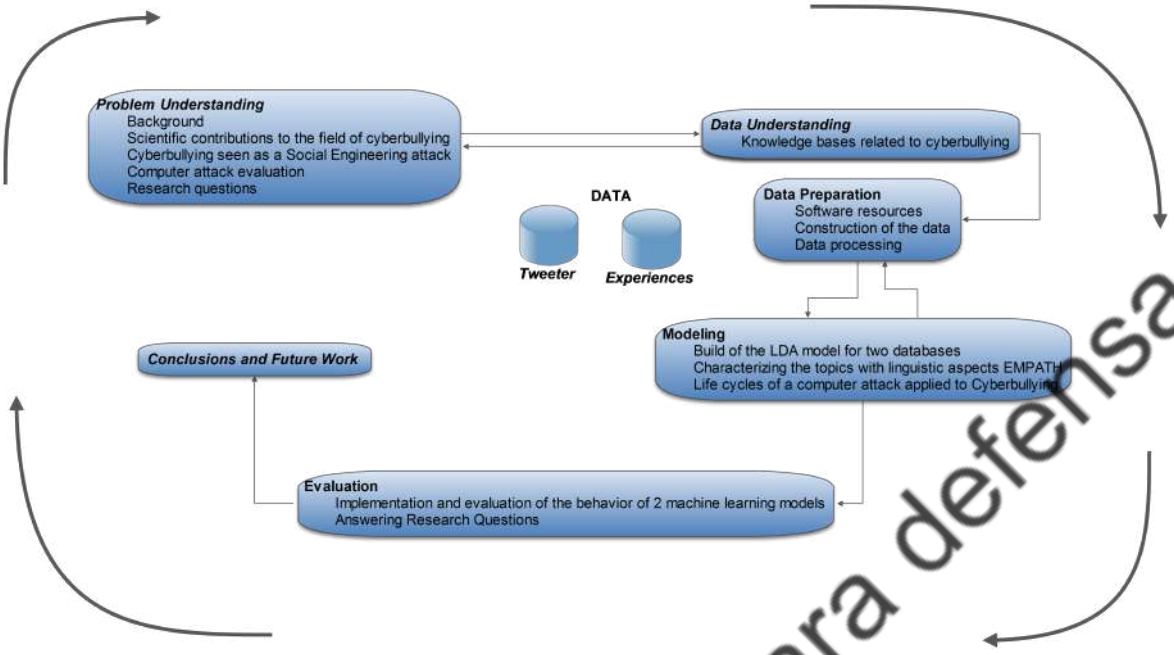


Figure 5.1: Methodology followed in this study based on [2].

formation security is evaluated. Finally, the main objectives and relevant research questions will be introduced.

5.4.1 Background

Cyberbullying is an attack that shows aggressive and intimidating behavior directed mainly at children and adolescents by attackers. Research on this topic demonstrates the existence of short-, medium-, and long-term effects on the emotional and social development of victims [3]. Their vulnerability characterizes these types of victims, and it is exploited by the attackers to exercise social dominance and abuse of power over them [4].

Cyberbullying remains a common problem in electronic societies, especially among young people and within organizations. It can cause social, cultural, technical, financial, and, critically, psychological problems for victims [5]. This phenomenon is considered the most dangerous type of attack since it represents a risk to users'social security in social networks, so researchers have studied how to mitigate these risks, threats, and problems caused by humans from the social network [6]. Besides, cyberbullying can result in cybercrime that violates existing laws as it happened in 2011; two girls, one of 11 and the other of 12, were accused of cyberbullying and computer intrusion in first grade for the crimes they allegedly committed against another 12-year-old girl, who was identified as a former friend. The couple

was accused of posting sexually explicit photos and messages on the victim's Facebook profile after obtaining their password [7].

5.4.2 Scientific contributions to the field of cyberbullying

In [7], the authors consider cyberbullying as an unconventional computer attack since the attacks are directed towards a specific victim (that can be a child or adolescent) through messages or comments of harassment. This attack is often evident in social networks where victims suffer intimidation, harassment, humiliation, and even threats [8]. Table 6.1 presents a literature review concerning the cyberbullying field.

In what follows, the most relevant contributions that support this research are summarized from the literature review described above.

- Detection and classification of cyberbullying attackers and victims using sentiment analysis techniques [9, 35].
- Approach of a computational model for the detection of cyberbullying through experimentation, with the use of mathematical algorithms, simulations and classification of behavior patterns [11].
- Identification and classification of the severity of words by the stalker, using support vector machine (SVM) classifiers [12].
- Development of an agent-based model (ABM), which analyzes the interactions occurring between the victim and the stalker [17].
- Cyberbullying detection using computer learning and expert systems by assigning a score that indicates the level of "bullying.ºf online attackers. For these investigations, knowledge bases from different social networks were used [33, 30].
- Prediction of the pattern of stalkers'writing in order to detect and prevent bullying attacks, using neural networks and SVM classifiers [18].
- Creating dictionaries of pejorative terms related to cyberbullying that are used in social networks, using text mining techniques and natural language processing (NLP) [37, 63].

Table 5.1: Cyberbullying - Literature review

Literature review	***Reference***
Classify stalkers and cyberbullying victims in a graphic model	[9]
Determine the correlation based on age and gender of cyberbullying victims	[10]
Development of a computational model that analyzes patterns of texts sent by the stalker	[11]
Label the texts according to their severity using the technique of (SVM)	[12], [13]
Label the information to relate it to a series of characteristics such as cyber aggression, cyber bullying by analyzing the behavior of comments directed towards the victim	[14]
Analyze the behavior of users in different line environments for detection and prevention, as well as the context of their texts with machine learning techniques, Multinomial Naïve Bayes, SVM, tree of decision	[15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29]
Determine levels of intimidation of stalkers towards victims	[30]
Assess the risk of cyberbullying with the use of TIC	[31], [32]
Develop models based on machine learning to transfer to other data sets	[33]
Development of a form with questions about online cyberbullying	[34]
Detect cyberbullying by feeling analysis	[35], [36], [28], [37] , [38]
Analyze comments from the two users to determine conversation patterns, using neural networks, brute force algorithms	[6], [39], [29]
Increase training data and identify offensive words with data mining techniques	[40], [41], [37], [42], [43]
Detect cyberbullying with machine learning, machine learning, NLP	[44], [45], [46], [47], [48], [49], [50], [51]
Classify comments with sentiment analysis, NLP and data mining techniques	[38], [52], [53], [54], [55], [56], [57], [58]
Sort comments to get word clouds	[59]
Identify and measure the degree of cyberbullying in social networks	[60]
Determine cyberbullying qualitatively in three classes in the Arabic language	[61]
LDA to have topics that are associated with a set of discussion topics	[62]
Analyze algorithm and methods to say which is the best in the case of cyberbullying	[51]

- Analyze social networks like Facebook or Twitter to generate a weighted function score of the texts related to cyberbullying and generate topics for discussion with the use of the latent Dirichlet (LDA) allocation algorithm [61, 62].

- Detection cyberbullying with machine learning methods for optimal learning in order to detect cyberbullying in social networks and prevent the victim [44, 45, 46, 47, 48, 49,

50].

- ❖ Analyzing text mining algorithms and methods with the purpose of which technique is appropriate to analyze cyberbullying [51].

As a relevant aspect, it is worth noting that, based on scientific evidence, the research conducted in this study has not been presented in similar works, thus generating new contributions to the field.

5.4.3 Cyberbullying seen as a Social Engineering attack

The identification and classification of threats are the basis for building defense mechanisms. Humans are considered the weakest link in information security due to their susceptibility to different manipulation techniques [64]. Concepts associated with "social engineering," define it as a dark art in computer science where deception is the primary tool to induce a person to divulge private information involuntarily and, in this way, attackers gain access to computer systems [65]. Kevin Mitnick said it is much easier to trick someone into presenting his credentials than using sophisticated piracy maneuvers [66]. The field of social engineering is still in its early stages of standardization with formal concepts, frameworks, and work templates [67]. The *art* of influencing people to divulge sensitive information is known as social engineering, and the process of doing it is known as social engineering attack. Nowadays, there are definitions of social engineering and several different models of social engineering attacks [64].

Based on a previous study [68], we performed an analysis to establish whether a cyberbullying attack is associated or not with social engineering. The analysis is performed following the four stages proposed by Kevin Mitnick [67], namely information gathering, development of relationship, exploitation of relationship, and execution to achieve the objective. We extract the relevant data from the selected literature and match it to the four stages mentioned earlier. The main findings are summarized in Table 6.2 and described as follows.
(1) *Information gathering.* This stage consists of gathering as much information as possible about the objective. This information is used in the following stages and is vital for the quality of the result in cases of targeted attacks. In the study of cyberbullying and its knowledge bases, it is evident that in most cases, the attackers have prior knowledge of their victims concerning their religion, gender, and physical characteristics.

(2) *Development of relationship.* In some experiences related to cyberbullying, it is evident that before the execution of the attacks, there were previous communications where the attackers and their victims were related in trusted environments. In several cases, these communications were made using websites, social networks, blogs, chat rooms (instant messaging), and email.
(3) *Exploitation of the relationship.* At this stage, the more significant activity related to cyberbullying is evident, given that by having direct contact with the victim and knowing their vulnerabilities, the attackers exert more pressure; as a result, their victims are affected, thus achieving isolation, frustration, low self-esteem, the depression between other effects. It cannot be evidenced that the attacker follows elaborate strategies for this purpose; however, the typical pattern of all attackers is to upset their victims psychologically.
(4) *Execution to achieve the objective.* The investigations related to this field are still inconclusive about the purpose or objective of the attacks; however, analyzing the data, aspects related to violence, physical and psychological damages, and suicide is evidenced in some cases.

From the previous evaluation, we can determine that the cyberbullying attack is associated to the field of social engineering.

5.4.4 Computer attack evaluation

A computer attack represents any hostile activity against a system or a person, using computer applications or psychological persuasion techniques. Every attack has a target, and scientists'responsibility is to determine what they are in order to apply defense strategies. It is worth noting that computer attackers are aware of the development and execution of each attack by developing a series of phases, stages, or steps to follow to make an attack successful [69, 70, 71, 72].

To identify the effects of each attack, they should be interpreted not only as isolated incidents or intrusions but also as operations that, in some cases, contemplate long periods. The stages of a computer attack are represented by models of life cycle applicable to cyberattacks, as illustrated in Figure 5.2; they are also known as "cyberattack chains" [73, 74, 75]. In the scientific field, several authors refer to the life cycle approach of Lockheed Martin [76], who developed an initial cyberattack chain model.
Under this criterion, this study's main contribution is the theoretical/practical definition of the

Table 5.2: Scientific evidence that relates cyberbullying to social engineering

Reference	Information gathering	Development of relationship	Exploitation of relationship	Execution to achieve the objective
Sentiment Analysis for Effective Detection of Cyber Bullying[9]			***	
Computational Analysis of Bullying Behavior in the Social Media Era[11]			***	
Time Series Forecasting in Cyberbullying Data[18]	***		***	***
Investigating Japanese Ijime (Bullying) behavior using agent-based and system dynamics models[17]		***	***	
Experts and Machines Against Bullies: A Hybrid Approach to Detect Cyberbullies[30]			***	
Sentiment Informed Cyberbullying Detection in Social Media[35]	***	***	***	
Detection of Behavior Patterns through Social Networks like Twitter, using Data Mining techniques as a method to detect Cyberbullying[37]			***	***
Exploratory Research to Identify the Characteristics of Cyber Victims on Social Media in New Zealand[10]	***	***	***	
Improving Cyberbullying Detection with User Context[16]	***	***		
BullyBlocker: Towards the identification of cyberbullying in social networking sites[60]			***	***

life cycle of bullying, from the perspective of information security. We elaborate on this matter in Section 5.7.3.

5.4.5 Research questions

Formalizing the concept of cyberbullying within the field of information security will allow researchers to support future research related to social engineering with the contributions generated in this field. To achieve these objectives, we formulated the following research questions:

RQ1: Can the data obtained by victims and attackers be used to define a functional model

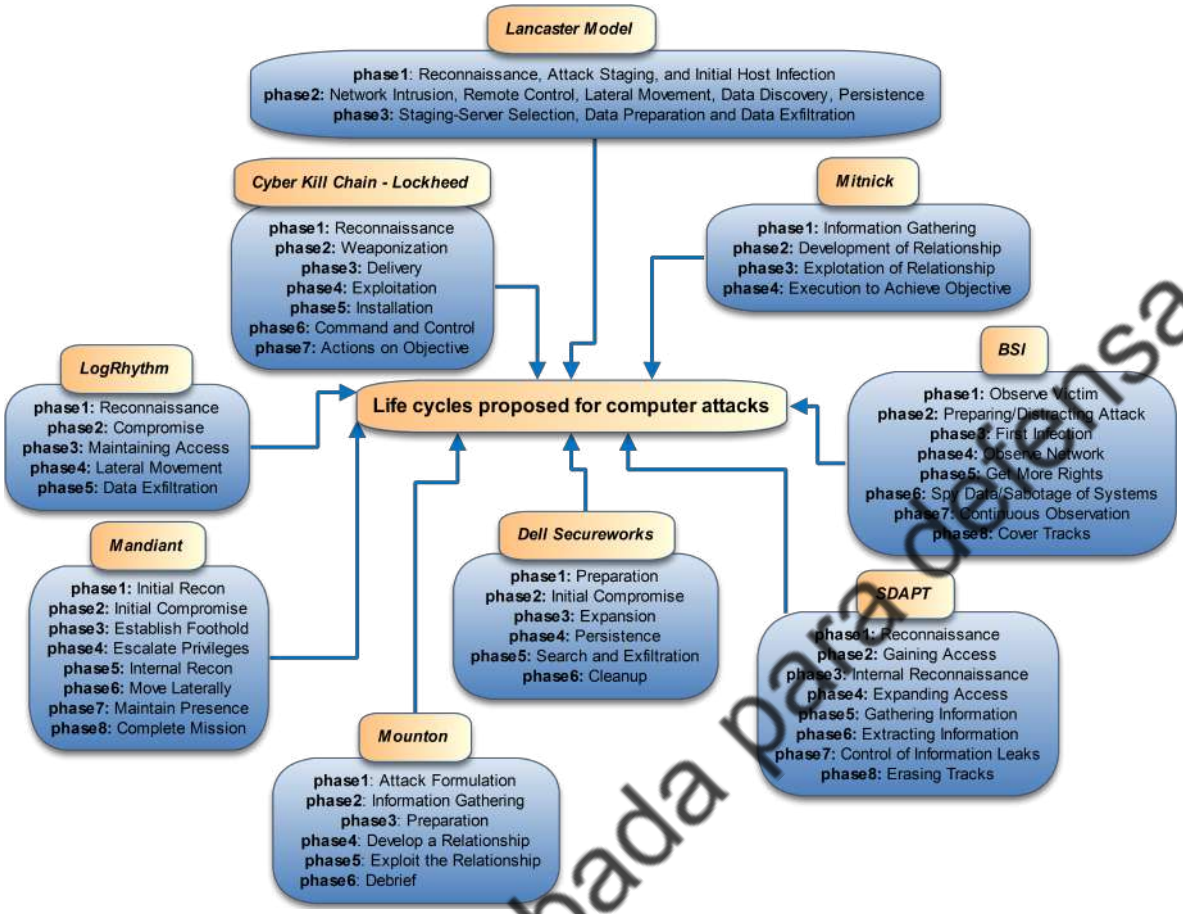


Figure 5.2: Life cycles proposed for computer attacks [72, 67].

for bullying?

RQ2: Can the phases of cyberbullying be determined as a computer attack by using topic modeling?

RQ3: Can the studies related to bullying support future research associated with social engineering?

5.5 DATA UNDERSTANDING

5.5.1 Knowledge bases related to cyberbullying

In this phase, two activities were carried out. The first activity was to evaluate and to select processes or procedures for obtaining data related to cyberbullying on social networks. The second activity was to assess and download data from websites that contain information related to the experiences of victims of these attacks, as well as techniques for downloa-

ding data. These activities allow us to find common aspects and understand the nature of cyberbullying to model it.

The research analyzes two types of information sources related to cyberbullying. Since the issue has been addressed only from the attackers'point of view, the investigation embraced a procedure for downloading tweets proposed by Lee et al. [77], thus improving the process of filtering and downloading the information. On the other hand, as there was no scientific evidence that addressed cyberbullying victimséxperiences, several websites were analyzed (see Table 5.3). Based on their structure and information quality, we proceeded to select the websites `https://pacerteensagainstbullying.org/` and `https://pacerkidsagainstbullying.org/` as knowledge bases. It is worth mentioning that the selected websites are created by the PACER National Center for Bullying Prevention (NBPC). Since 2006, PACER NBPC has actively led social change to prevent child bullying. As additional data, it is highlighted that all the experiences are independent, and the loading process is not directed, nor do they follow standard guidelines that may bias the data in our results. The technique used to download the experiences is Scrapy.

In the structural analysis of the data to be processed, it can be seen that the messages corresponding to tweets have a typical structure; it consists of five parts: the name of the user, the date, the links to the retweets, the message, and the hashtags. On the other hand, the experience texts maintain a homogeneous structure; in this context, three parts were determined: user, date, and message. From the two knowledge bases, it was determined that the essential part of the analysis was the message. The remaining parts were considered as not relevant to the present study.

5.6 DATA PREPARATION

In the following, the tools and processes used for debugging and downloading information is described.

5.6.1 Software resources.

An essential aspect to consider on the scientific method is reproducibility. Under this criterion, Table 6.3 describes the software resources and databases used in the research.

Table 5.3: Cyberbullying Related Websites

http://www.athinline.org/overtheline
https://pacerteensagainstbullying.org/advocacy-for-others/real-teens-speak-out/
https://www.childhelplineinternational.org/data-overview/voices/
https://www.childline.org.uk/get-involved/real-life-stories/bullying-lolas-story/
https://www.ranker.com/list/worst-high-school-bully-horror-stories/jessika-gilbert
https://www.noplace4hate.org/real-bullying-stories/
https://www.glaad.org/amp/lgbtq-youth-share-stories-offer-advice-adults-to-end-bullying
https://www.quora.com/What-is-your-individual-story-about-escaping-bullying
https://www.news.com.au/lifestyle/relationships/six-australians-share-their-bully-stories-8216i-lost-faith8217/news-story/0aa0582b8d446cee966b9cf6330428fd
https://www.familyorbit.com/blog/real-life-cyberbullying-horror-stories/
https://www.thejournal.ie/readme/bullied-your-stories-of-bullying-and-its-lasting-effects-665141-Mar2013/
https://www.childline.org.uk/get-involved/real-life-stories/bullying-lolas-story/
https://crossingenres.com/a-short-story-on-bullying-a35a924877bd
https://www.unicef.org/eca/stories/shezanas-story-being-bullied-ending-conflicts-school
https://learnenglishkids.britishcouncil.org/sites/kids/files/attachment/video-zone-abbies-bullying-story-transcript.pdf
https://www.rd.com/true-stories/inspiring/bullying-in-school/
https://www.lolwot.com/10-real-stories-of-bullying-that-will-shock-you/
https://www.shepellfgi.com/Bullying/StoriesofBullying/US/83
http://eyesonbullying.org/will.html
https://kidshelpphone.ca/get-involved/open-door-carrie-story/
https://childmind.org/article/a-cyber-bullying-story-with-a-twist/
https://www.bbc.com/future/story/20190913-why-some-children-become-merciless-bullies
https://www.jackpetcheyfoundation.org.uk/our-stories/blogs/my-bullying-story-and-how-i-overcame-it/
http://www.buildfuture-stopbullying.eu/references/stories/
https://www.teenvogue.com/story/20-personal-bullying-stories
https://www.greatschools.org/gk/articles/combatting-bullying-or-is-this-just-the-new-normal/
https://soapboxie.com/social-issues/Being-Bullied-in-School-My-Own-Experience
https://tylerclementi.org/tylers-story/
https://kidshealth.org/en/parents/cyberbullying.html?WT.ac=p-ra

Table 5.4: Materials Used in Experimentation

Software Resources	
Resource	***Version***
Windows	N/A
MATLAB	R2019a Version 9.6
Text Analytics Toolbox	R2019a Version 9.6
Microsoft Excel	Excel 2017 Version 15.0
Anaconda	Anaconda 2019.10 Version 3.7
RapidMiner	RapidMiner Studio 9.6
R-Studio	RStudio Desktop 1.2.5033
Data Resources	
Resource	***Version***
hatebase	https:// hatebase.org
pacerteensagainstbullying	https:// pacerteensagainstbullying.org/
pacerkidsagainstbullying	https:// pacerkidsagainstbullying.org/

5.6.2 Data construction.

This phase describes all the activities necessary for the data construction to be processed from the initial raw data.

- Activity 1. *Data selection*. The data sets to be downloaded are comprised of 250,000 attacker-related tweets and 3,035 victim experiences. Then, the properties of these are determined, such as the number of message lines contained in each conversation, the number of words contained, verification of linguistic structures. It should be noted that determining the quality of the data acquired was a challenge because the texts come from various sources, use informal language, the vocabulary consists of slang, shorthand, emoticons and contain spelling errors.

- Activity 2. *Downloading the data*. The investigation established that the file format should be CSV. To download the experiences automatically in CSV format, we used the Scrapy technique (see Figure 5.3) by developing a Python script. In the script, the extraction website URLs and rules were defined to access the rest of the experiences

automatically. For this purpose, the Xpath language was used.

Furthermore, a seed-based tweet extraction model was performed using Rapidminer and R studio. Within the process of downloading tweets (see Figure 5.4), 250 keywords ("seed") from `hatebase.org` were considered following the procedure in [77]. Then, the obtained data will go through information purification processes. These processes are stemming, normalization, conversion of text to lowercase letters, elimination of punctuation marks, stopwords, words with two or fewer characters, and words with 15 characters or more.

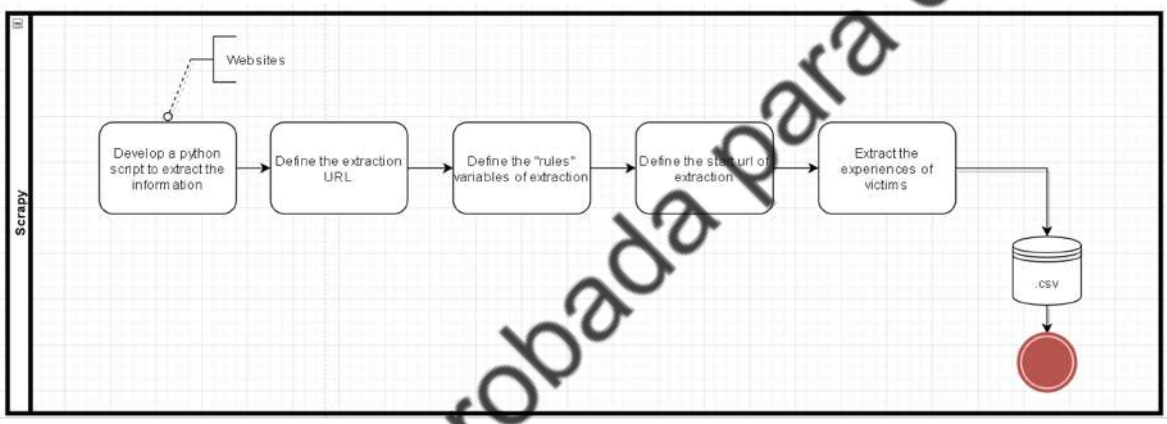


Figure 5.3: Scrapy technique.

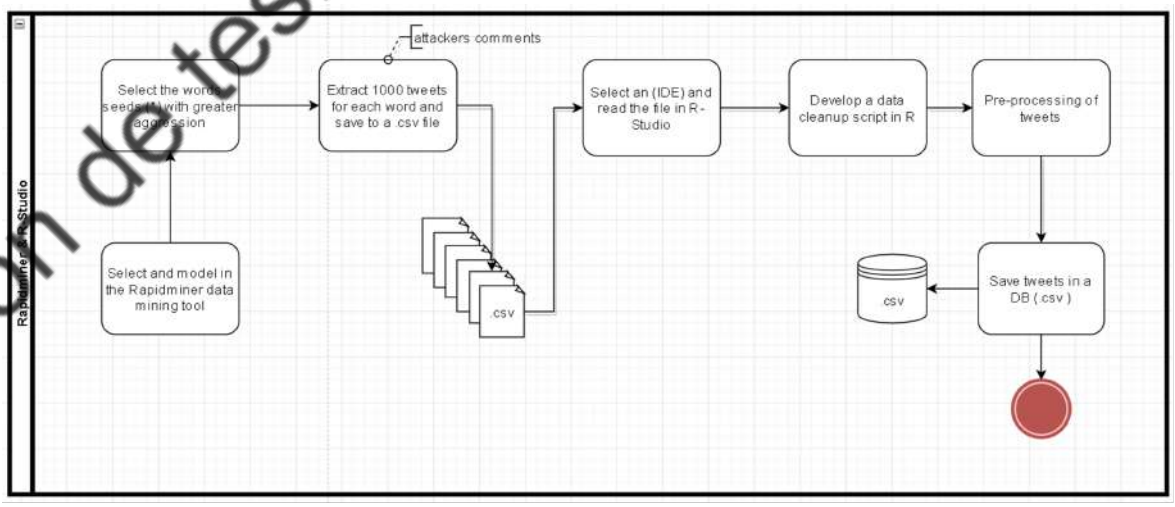


Figure 5.4: Rapidminer and R Studio.

5.6.3 Data processing.

For data processing, we developed a Matlab function (see Figure 5.5) that performs two intermediate threads called cleaning and text standardization. Cleaning the text consists of eliminating punctuation marks, special characters, and words that add noise to the study. These words are known as stop words and are all those articles, prepositions, conjunctions, pronouns, among others, that do not add meaning to the investigation. After cleaning, the standardization stage is performed where all the text is lowercase; the verbs are taken to their base form, for example “getting to get,” lemmatization and normalization techniques are applied, to finally eliminate all words that have two or fewer characters or that exceed 15 characters. Figure 5.6 shows the difference between the original and processed data.

```
function [documents] = preprocessText(textData)
% Convert the text data to lowercase.
cleanTextData = lower(textData);

% Tokenize the text.
documents = tokenizedDocument(cleanTextData);

% Erase punctuation.
documents = erasePunctuation(documents);

% Remove a list of stop words.
documents = removeStopWords(documents);

% Remove words with 2 or fewer characters, and words with 15 or greater
% characters.
documents = removeShortWords(documents,2);
documents = removeLongWords(documents,15);

% Lemmatize the words.
documents = addPartOfSpeechDetails(documents);
documents = normalizeWords(documents,'Style','lemma');
```

Figure 5.5: Function developed in Matlab for data preprocessing.

5.7 MODELING

Topic modeling belongs to the field of text mining. It allows for analyzing many unstructured texts. There are several topic modeling methods, among the most proven are Latent Semantic Analysis (LSA), Probable Latent Semantic Analysis (PLSA), Correlated Theme Model (CTM), and LDA. The latter has been evaluated and compared to other topic modeling algorithms. Criteria such as “LSI or LSA is much faster to train than LDA, but has less precision,” “LDA can describe the real semantics,” “LDA has all the inherent skills of PLSA;

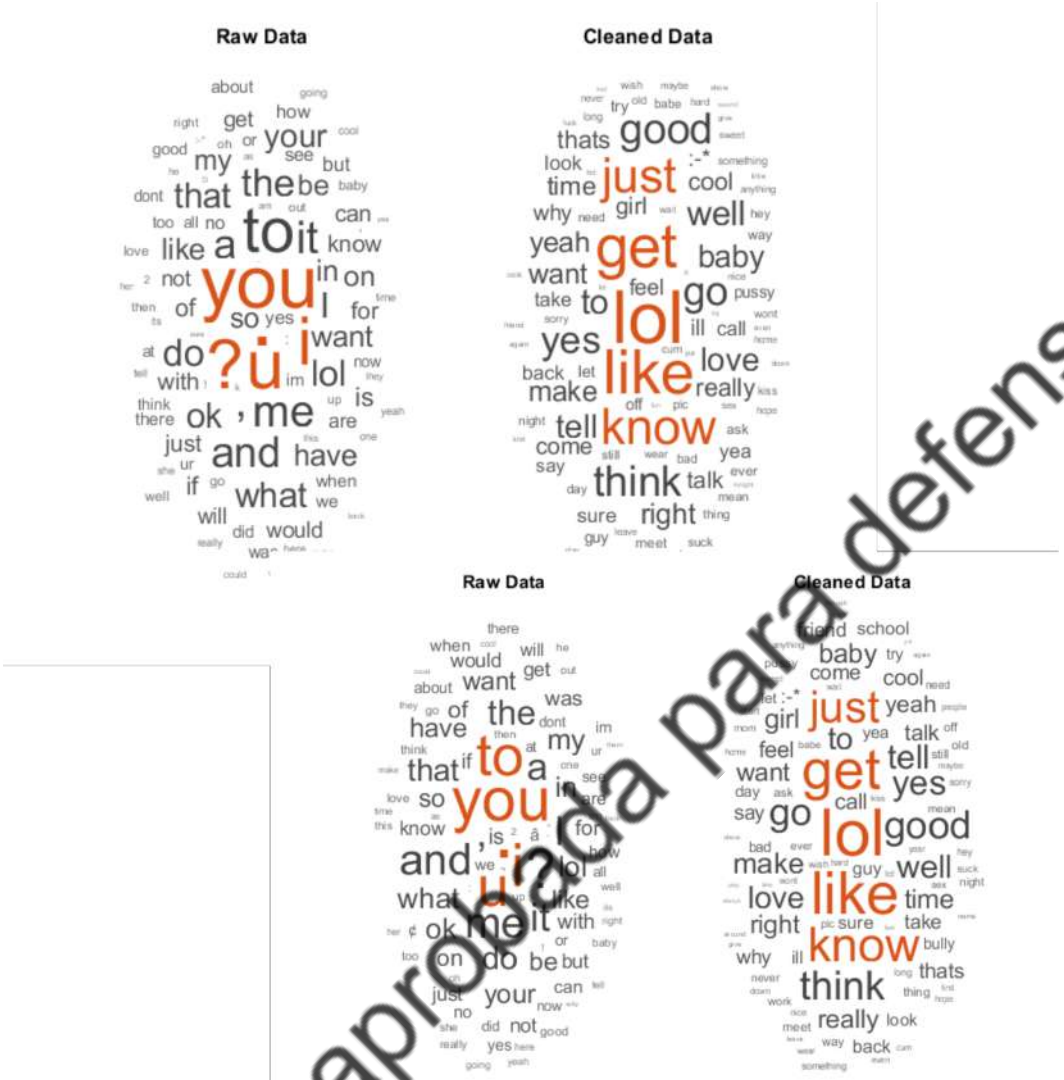


Figure 5.6: Word cloud of pre-processed (raw data) and processed text (cleaned data). Tweets at top and experiences at the bottom.

however, its results are more close to real semantic environments," "LDA directly chooses a suitable set of themes from the theme distribution compared to PLSAs that need a prior probability," and "LDA works better than PLSAs because it can easily be generalized to new documents" [78, 79, 80, 81] position LDA as a leading model in the field of topic modeling.

5.7.1 Building the LDA model for two databases

In topic modeling, the LDA approach considers each document as a collection of topics in a particular proportion. Besides, each topic is considered as a collection of keywords. For the implementation of the model, two parameters are required basically: (i) the corpus of

information previously cleaned, and (ii) the number of topics.

5.7.1.1 Perplexity assessment to determine the optimal number of topics

Using the refined database, we will proceed to determine an appropriate number of topics. It is worth mentioning that the LDA model can be evaluated based on perplexity and processing time. For this purpose, a function was developed to visualize the behavior of the model with various themes, thus determining an optimal compromise (cycles that will determine the model of cyberbullying). The results of this analysis are presented in Figures 5.7 and 5.8. After the perplexity analysis, four topics were established as optimal values for applying the LDA model. With the number of topics and the knowledge bases, we proceeded to build the LDA model applicable to Cyberbullying.

5.7.1.2 Building the LDA model applicable to Cyberbullying

Figures 5.9 and 5.10 describe the LDA model built with four topics and the pre-processed databases (tweets and experiences). Each topic is made up of keywords combination, which in turn gives relevance to the topic through their weights. As can be seen in the previous figure, the model fits better with the data of the experiences and does not give good results with the tweets. This is because the tweet messages, by their nature, are concise and do not adapt to the LDA model. However, the experiences knowledge base classified the words without overlaps in the four topics mentioned earlier.

Although the tweets'database was not adapted to the LDA model to evaluate the phases of the attack, it will be useful to create dictionaries that relate to the linguistic aspects raised in each phase; this process was developed to implement computational intelligence algorithms.

5.7.2 Characterizing the topics with linguistic aspects using EMPATH

After obtaining the optimal LDA model with four topics, we proceeded to give a linguistic meaning to each group of words (equivalent to dictionaries). To develop this process, an online application called EMPATH [82] was used. The process is described in Figure 5.11,

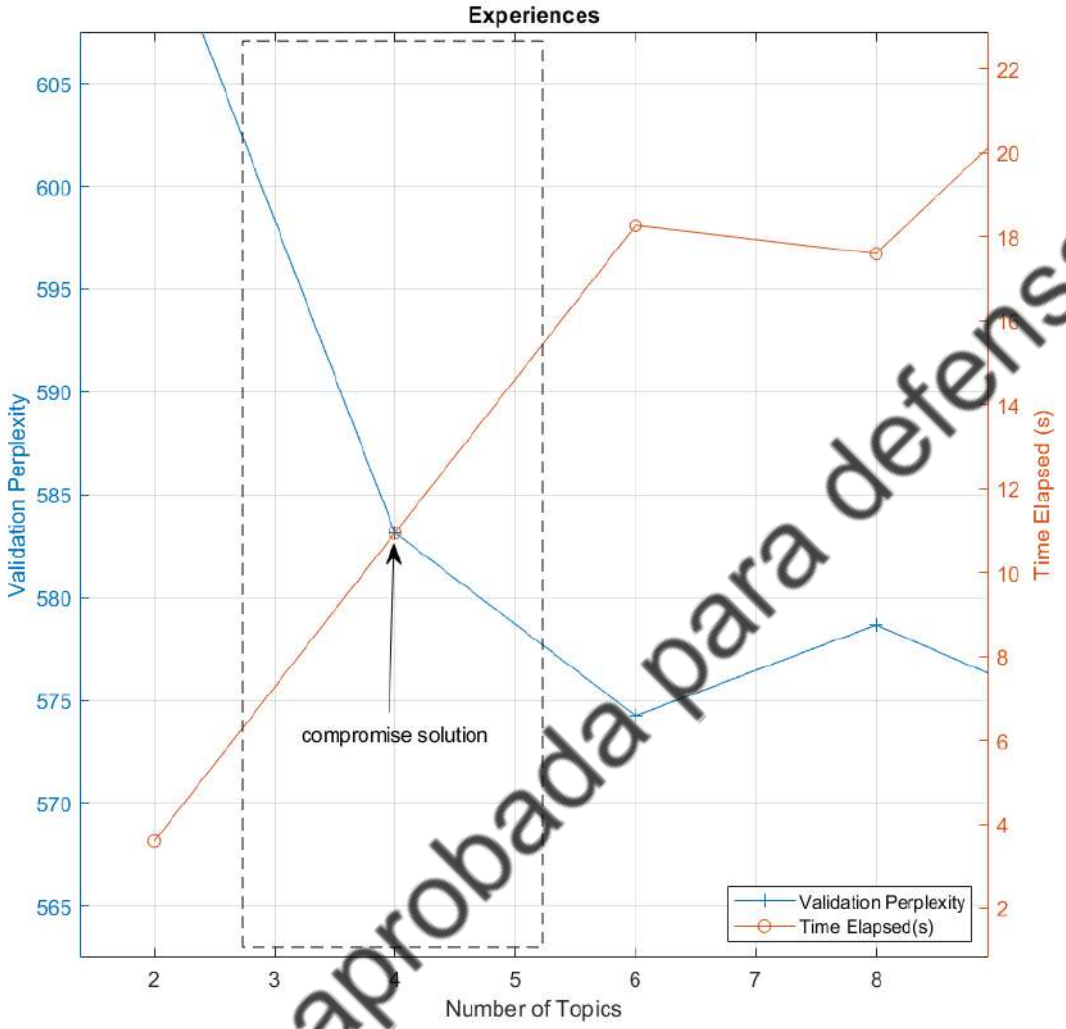


Figure 5.7: Topics assigned to cyberbullying by LDA - Data Experiences.

the most important words from each word bag are the inputs, and the application returns lexical categories to create a unique concept that identifies each topic. It should be noted that this application takes each word from the given set and looks for similarities between the words that define a category; then, the application validates which category the word belongs to among the 200 existing ones. For doing so, artificial intelligence based on neural networks is used. The lexical categories obtained for each topic are:

- Topic 1.- friends, children, party, family, positive_emotion happiness, youth and feminine;
- Topic 2.- Home, pain, negative_emotion, violence, shame, and body;
- Topic 3.- Communication, speaking, phone, listen and hearing;

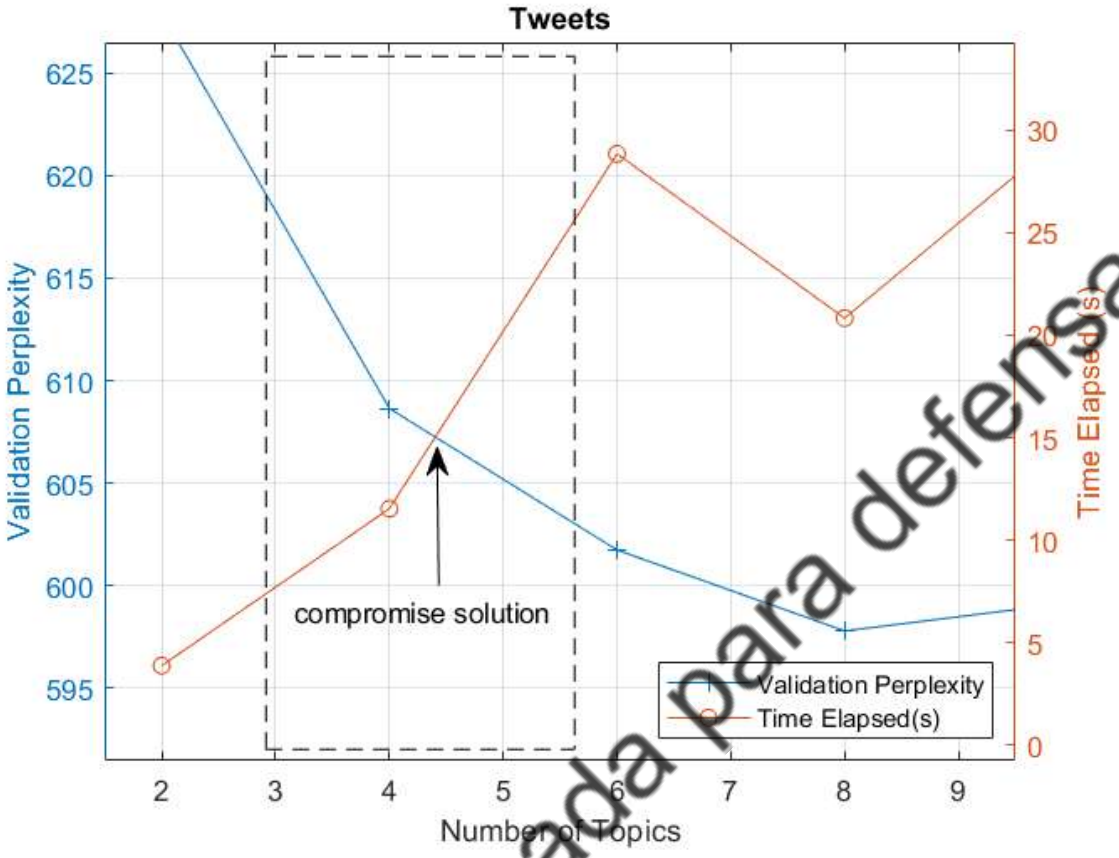


Figure 5.8: Topics assigned to cyberbullying by LDA - Data Tweets.

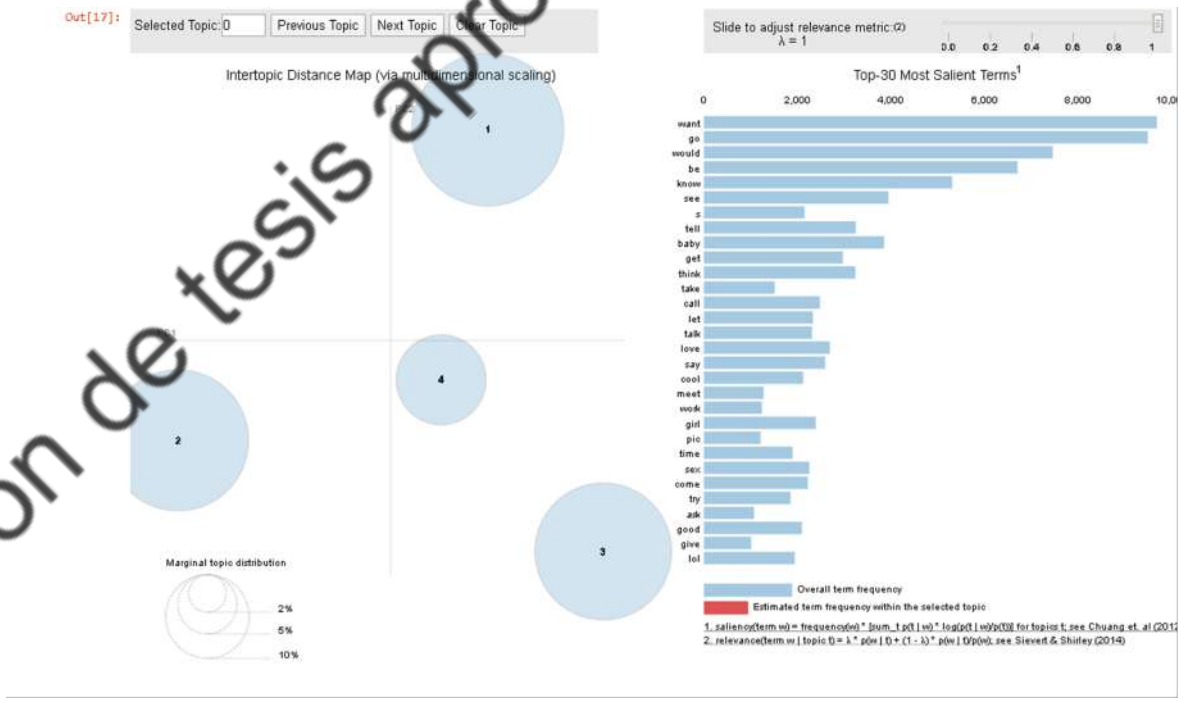


Figure 5.9: Model building with Experiences.

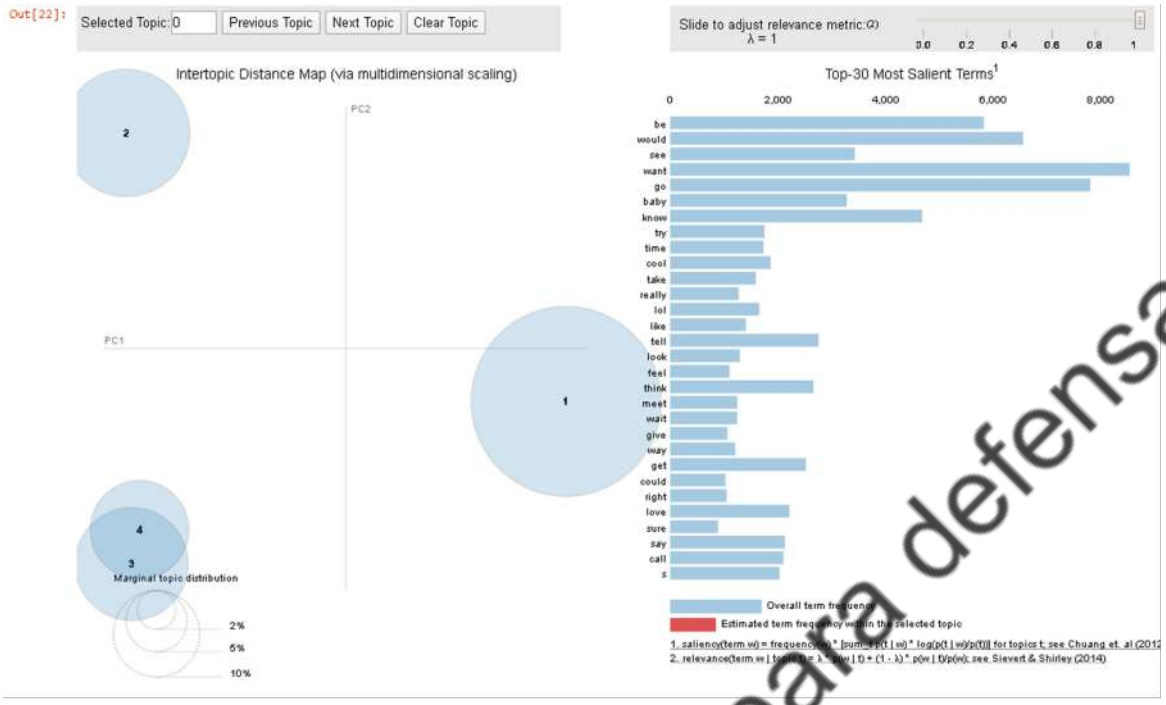


Figure 5.10: Model building with Tweets.

❖ Topic 4.- Leisure, movement, sexual, feminine, body, work and childish.

5.7.3 Life cycles of a computer attack applied to cyberbullying

To analyze cyberbullying as a computer attack, we proceeded to verify the relationship of the most representative life cycles with the linguistic descriptors described in Figure 5.12.

After this process, we proceeded to determine a specific definition of each topic, comparing them conceptually with the proposed phases within the life cycles associated with computer attacks.

❖ In the analysis of the first topic, the conceptual correlation with several stages was evident; however, the concept that most adapted to the linguistic description was **Development of Relationship**, of the Mounton model.

❖ Regarding the second topic, the definition most related to linguistic descriptors was **Preparing/Distracting Attack** of the BSI model.

❖ Concerning the third station, its linguistic descriptors pointed to the **exploit the relationship**, a concept found in the Mounton model.

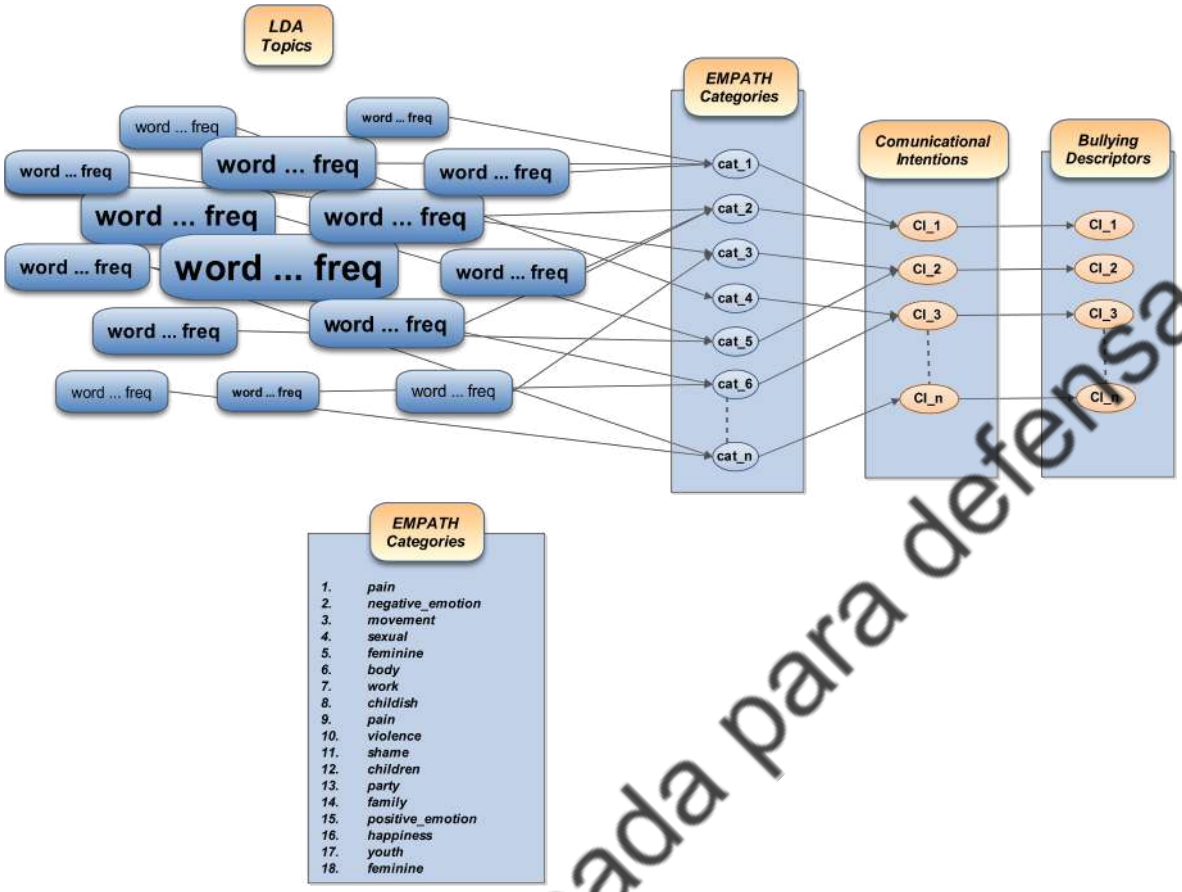


Figure 5.11: EMPATH operation.

- For the selection of the last station concept, it was observed that almost all the models'stations were contrasted with the last or penultimate station. However, the phase proposed by Mounton et al. [83] is the best one that describes the end of the attack **Debrief**, since in this station, the attacker manages the mental state of his victim at his convenience with different strategies.

5.8 EVALUATION

In this phase of the investigation, two activities will be conducted: (i) implementing and evaluating the behavior of two machine learning models to identify their precision with new data, and (ii) evaluating the proposed model to answer the research questions.

Lockheed	1	Reconnaissance				
	2	Weaponization				
	3	Delivery				
	4	Exploitation				
	5	Installation				
	6	Command & Control (C2)				
	7	Actions on Objective				
LogRhythm	1	Reconnaissance				
	2	Compromise				
	3	Maintaining Access				
	4	Lateral Movement				
	5	Data exfiltration				
Mandiant	1	Initial Recon				
	2	Initial Compromise				
	3	Establish Foothold				
	4	Escalate Privileges				
	5	Internal Recon				
	6	Move Laterally				
	7	Maintain Presence				
	8	Complete Mission				
Dell Secureworks	1	Preparation				
	2	Initial Compromise				
	3	Expansion				
	4	Persistence				
	5	Search and Exfiltration				
	6	Cleanup				
SDAPT	1	Reconnaissance				
	2	Gaining access				
	3	Internal Reconnaissance				
	4	Expanding Access				
	5	Gathering Information				
	6	Extracting information				
	7	Control of Information leaks				
	8	Erasing Tracks				
BSI	1	Observe Victim				
	2	**Preparing/Distracting Attack**		***		
	3	First Infection				
	4	Observe Network				
	5	Get More Rights				
	6	Spy Data/Sabotage of Systems				
	7	Continuous Observation				
	8	Cover Tracks				
Mitnick	1	Information Gathering				
	2	**Development of Relationship**	***			
	3	Exploitation of Relationship				
	4	Execution to Achieve Objective				
Lancaster	1	Reconnaissance, Attack Staging, and Initial Host Infection				
	2	Network intrution, Remote Control, Lateral Movement, Data Discovery, Persistence				
	3	Staging-Server Selection, Data Preparation and Data Exfiltration				
Mouton	1	Attack Formulation				
	2	Information Gathering				
	3	Preparation				
	4	Develop a Relationship				
	5	**Exploit the Relationship**			***	
	6	**Debrief**				***
			Development of Relationship	**Preparing/Distracting Attack**	**Exploit the Relationship**	**Debrief**

Figure 5.12: Operational concepts (definition) related to the linguistic characteristics.

5.8.1 Implementing and evaluating the behavior of two machine learning models

Once we defined the cyberbullying attack model and its stages, we implement and evaluate the precision of two learning models with new data. In selecting the technique for classifying the supervised learning model, the characteristics of cyberbullying were considered. Under this background, a linear classifier was selected. It is characterized by the simplicity of implementation and computational appeal. It is worth mentioning that while linear classifiers are not designed to work with complex data, they may be sufficient for textual content data. For example, linear classifiers have been shown to offer competitive returns on document data with nonlinear classifiers. A significant advantage of linear classification is that the training and evaluation procedures are much more efficient. Therefore, linear classification can be beneficial for some large-scale scenarios[84, 85, 68].
On the other hand, a Long-term Memory (LSTM) network was implemented. This type of network, also known as a recurring network, is characterized by storing or recalling previous information states and using them to decide what the next state will be [68].

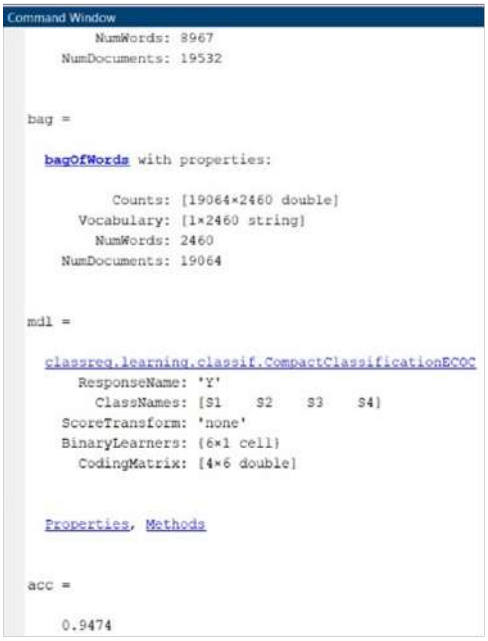
5.8.1.1 Data processing and results

For implementing the two models, the knowledge base was divided into two groups. The first group, called *training* contains 70 % of the data that will serve for the training of the models and the second group called *test* contains 30 % of the data that will be used to check the precision of the model at the time of processing unknown information.

The classification accuracies obtained with these models were 94.74 % for the linear classification and 94.41 % for the LSTM network, as illustrated in Figure 5.13. Based on these results, it is shown that the linear model fits best with our case study since its precision is higher (94.74 %) than the deep learning model.

These results show that the model is highly applicable to an early detection system for unconventional attacks.

Linear algorithm response



LSTM response

Figure 5.13: Results of the implementation of two machine learning algorithms.

5.8.2 Answering the research questions

5.8.2.1 Can the data obtained by victims and attackers be used to define a functional model for bullying?

In this case study, it was evidenced that the tweets'knowledge base obtained could not establish topics, unlike the knowledge base related to the experiences. This result is because the tweets contain very dispersed information, and this knowledge base does not adapt to the LDA model, contrary to the structured information of the experiences, which established four very well-defined topics. Under this criterion, the research question is answered.

5.8.2.2 Can the phases of cyberbullying be determined as a computer attack by using topic modeling?

Some studies related to cyberbullying have been conducted from the psychological perspective; this aspect not being supported by the scientific community's agreements allows researchers to determine different phases with a high degree of subjectivity. For this reason, this research was based on the LDA statistical model that allowed determining a specific number of stations or phases that attackers follow when applying this attack to their victims. With that knowledge, cyberbullying can be established as a computer attack.

5.8.2.3 Can the studies related to bullying support future research associated with social engineering?

Considering that social engineering studies are in continuous development and have one of their objectives to determine behavioral patterns of the attackers and their victims, it is evident that the present study can support future investigations aligned to the study of social engineering within the information security field.

5.9 CONCLUSIONS AND FUTURE WORK

Detection and mitigation mechanisms that counter unconventional attacks, where the means of communication is human interaction, are still under development. As evidenced in the literature review, efforts have been made to eliminate the effects of this phenomenon; however, it persists and manifests more severely in its victims. Studying this phenomenon from the information security point of view, such as social engineering, allows linking techniques, methodologies, and proposed architectures to cyberbullying and vice versa. With this link, it would be possible to develop more specialized knowledge on the subject, thus avoiding incomplete and scattered proposals. The results obtained in this investigation justify the validity of the process of obtaining unconventional attack models, as evidenced in a previous investigation related to cyberbullying.

We have positioned cyberbullying as an attack associated with social engineering and information security. Through the modeling of topics, different stages or seasons of a life cycle

of cyberbullying associated with social engineering is determined; this will allow supporting investigations related to identifying patterns of malicious behavior online. In the experimentation phase, a typical pattern of cyberbullies was determined in the processing of information related to the experiences of the victims. The LDA statistical and computational algorithm classified all the information into four topics or stations that were part of the cyberbullying model. By themselves, these classifications did not describe linguistic aspects; for this reason, linguistic software was used to define the communicational intentions of each stage. With this knowledge, the stages of the models attached to the information security field were correlated, and the definitive model of our case study was defined.

The research covers several fields aligned to security, such as social engineering. The processing and evaluation of short text lines obtained from instant messaging protocol, through the proposed approach, does not only apply to the case study but can be reproduced in other security-related fields, these can be online bullying, bank fraud, phishing, among others. One of the main challenges in the path of new cases of study is obtaining relevant data related to the research field; for this reason, it is essential to promulgate and disseminate in the scientific community this type of study to gain more interest in the academy and industry. As future work, a prototype of parental control is in the development stage that, in the subsequent will require optimization with real case studies. As a previous step to this implementation, the model must be contrasted with instant message data, with conversational texts of different nature. In this way, the system will have the ability to decouple these conversations from the classification and location of texts relevant to the violation of privacy.

REFERENCES

[1] Hendro Margono. *Analysis of the Indonesian Cyberbullying through Data Mining: The Effective Identification of Cyberbullying through Characteristics of Messages*. PhD thesis, Victoria University, 2019.

[2] Steffen Huber, Hajo Wiemer, Dorothea Schneider, and Steffen Ihlenfeldt. Dmme: Data mining methodology for engineering applications–a holistic extension to the crisp-dm model. *Procedia CIRP*, 79:403–408, 2019.

[3] Matthew C Aalsma and James R Brown. What is bullying? *Journal of Adolescent Health*, 43(2):101–102, 2008.

[4] Anthony A Volk, Andrew V Dane, and Zopito A Marini. What is bullying? a theoretical redefinition. *Developmental Review*, 34(4):327–343, 2014.

[5] Ali Al Mazari. Cyber-bullying taxonomies: Definition, forms, consequences and mitigation strategies. In *2013 5th International Conference on Computer Science and Information Technology*, pages 126–133. IEEE, 2013.

[6] Haoti Zhong, David J Miller, and Anna Squicciarini. Flexible inference for cyberbully incident detection. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 356–371. Springer, 2018.

[7] kaspersky. Cyberbullying: The cybercrime of the century. urlhttps://usa.kaspersky.com/resource-center/ preemptive-safety/cyberbullying-and-cybercrime, 2020.

[8] Nour E Oweis, Mohammad A Alrababa, Waseem G Oweis, Suhail S Owais, and Mohammad Alansari. A survey of internet security risk over social networks. In *2014 6th International Conference on Computer Science and Information Technology (CSIT)*, pages 1–4. IEEE, 2014.

[9] Vinita Nahar, Sayan Unankard, Xue Li, and Chaoyi Pang. Sentiment analysis for effective detection of cyber bullying. In *Asia-Pacific Web Conference*, pages 767–774. Springer, 2012.

[10] Varun Dhond, Shahper Richter, and Brad McKenna. Exploratory research to identify the characteristics of cyber victims on social media in new zealand. In *European, Mediterranean, and Middle Eastern Conference on Information Systems*, pages 193–210. Springer, 2018.

[11] Fakhra Jabeen and Jan Treur. Computational analysis of bullying behavior in the social media era. In *International Conference on Computational Collective Intelligence*, pages 192–205. Springer, 2018.

[12] Carmen R Sedano, Edson L Ursini, and Paulo S Martins. A bullying-severity identifier framework based on machine learning and fuzzy logic. In *International Conference on Artificial Intelligence and Soft Computing*, pages 315–324. Springer, 2017.

[13] Laura P Del Bosque and Sara Elena Garza. Aggressive text detection for cyberbullying. In *Mexican International Conference on Artificial Intelligence*, pages 221–232. Springer, 2014.

[14] Homa Hosseinmardi, Sabrina Arredondo Mattson, Rahat Ibn Rafiq, Richard Han, Qin Lv, and Shivakant Mishra. Analyzing labeled cyberbullying incidents on the instagram social network. In *International conference on social informatics*, pages 49–66. Springer, 2015.

[15] Maral Dadvar, Roeland Ordelman, Franciska de Jong, and Dolf Trieschnigg. Towards user modelling in the combat against cyberbullying. In *International Conference on Application of Natural Language to Information Systems*, pages 277–283. Springer, 2012.

[16] Maral Dadvar, Dolf Trieschnigg, Roeland Ordelman, and Franciska de Jong. Improving cyberbullying detection with user context. In *European Conference on Information Retrieval*, pages 693–696. Springer, 2013.

[17] Chaiwat Thawiworadilok, Mohsen Jafari Songhori, and Takao Terano. Investigating japanese ijime (bullying) behavior using agent-based and system dynamics models. In *Australasian Joint Conference on Artificial Intelligence*, pages 554–558. Springer, 2015.

[18] Nektaria Potha and Manolis Maragoudakis. Time series forecasting in cyberbullying data. In *International Conference on Engineering Applications of Neural Networks*, pages 289–303. Springer, 2015.

[19] Jiamou Liu, Yingying Tao, and Quan Bai. Towards exposing cyberstalkers in online social networks. In *Pacific Rim International Conference on Artificial Intelligence*, pages 763–770. Springer, 2016.

[20] Mohammed Ali Al-Garadi, Mohammad Rashid Hussain, Nawsher Khan, Ghulam Murtaza, Henry Friday Nweke, Ihsan Ali, Ghulam Mujtaba, Haruna Chiroma, Hasan Ali Khattak, and Abdullah Gani. Predicting cyberbullying on social media in the big data era using machine learning algorithms: Review of literature and open challenges. *IEEE Access*, 7:70701–70718, 2019.

[21] I-Hsien Ting, Wun Sheng Liou, Dario Liberona, Shyue-Liang Wang, and Giovanny Mauricio Tarazona Bermudez. Towards the detection of cyberbullying based on social network mining techniques. In *2017 International Conference on Behavioral, Economic, Socio-cultural Computing (BESC)*, pages 1–2. IEEE, 2017.

[22] Batoul Haidar, Maroun Chamoun, and Ahmed Serhrouchni. Multilingual cyberbullying detection system: Detecting cyberbullying in arabic content. In *2017 1st Cyber Security in Networking Conference (CSNet)*, pages 1–8. IEEE, 2017.

[23] Vijay Banerjee, Jui Telavane, Pooja Gaikwad, and Pallavi Vartak. Detection of cyberbullying using deep neural network. In *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*, pages 604–607. IEEE, 2019.

[24] Aishwarya Upadhyay, Akshay Chaudhari, Sarita Ghale, SS Pawar, et al. Detection and prevention measures for cyberbullying and online grooming. In *2017 International Conference on Inventive Systems and Control (ICISC)*, pages 1–4. IEEE, 2017.

[25] Sourabh Parime and Vaibhav Suri. Cyberbullying detection and prevention: data mining and psychological perspective. In *2014 International Conference on Circuits, Power and Computing Technologies [ICCPCT-2014]*, pages 1541–1547. IEEE, 2014.

[26] Shahnoor C Eshan and Mohammad S Hasan. An application of machine learning to detect abusive bengali text. In *2017 20th International Conference of Computer and Information Technology (ICCIT)*, pages 1–6. IEEE, 2017.

[27] Monirah A Al-Ajlan and Mourad Ykhlef. Optimized twitter cyberbullying detection based on deep learning. In *2018 21st Saudi Computer Society National Computer Conference (NCC)*, pages 1–5. IEEE, 2018.

[28] Sreeshma Mohan, Indu Valsaladevi, and Sabu M Thampi. "think before you post": A cognitive psychological approach for limiting bullying in social media. In *International Conference on Smart City and Informatization*, pages 418–434. Springer, 2019.

[29] Sani Muhamad Isa, Livia Ashianti, et al. Cyberbullying classification using text mining. In *2017 1st International Conference on Informatics and Computational Sciences (ICICoS)*, pages 241–246. IEEE, 2017.

[30] Maral Dadvar, Dolf Trieschnigg, and Franciska de Jong. Experts and machines against bullies: A hybrid approach to detect cyberbullies. In *Canadian Conference on Artificial Intelligence*, pages 275–281. Springer, 2014.

[31] Sebastian Wachs, Wilfried Schubarth, Andreas Seidel, and Elena Piskunova. Detecting and interfering in cyberbullying among young people (foundations and results of german case-study). In *International Conference on Digital Transformation and Global Society*, pages 277–285. Springer, 2018.

[32] Jooyeon Won and DongBack Seo. Relationship between self-disclosure and cyberbullying on snss. In *European, Mediterranean, and Middle Eastern Conference on Information Systems*, pages 154–172. Springer, 2017.

[33] Sweta Agrawal and Amit Awekar. Deep learning for detecting cyberbullying across multiple social media platforms. In *European Conference on Information Retrieval*, pages 141–153. Springer, 2018.

[34] Himel Dev, Mohammed Eunus Ali, Jalal Mahmud, Tanmoy Sen, Madhusudan Basak, and Rajshakhar Paul. A real-time crowd-powered testbed for content assessment of potential social media posts. In *International Conference on Social Informatics*, pages 136–152. Springer, 2015.

[35] Harsh Dani, Jundong Li, and Huan Liu. Sentiment informed cyberbullying detection in social media. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 52–67. Springer, 2017.

[36] Mifta Sintaha and Moin Mostakim. An empirical study and analysis of the machine learning algorithms used in detecting cyberbullying in social media. In *2018 21st International Conference of Computer and Information Technology (ICCIT)*, pages 1–6. IEEE, 2018.

[37] Freddy Tapia, Cristina Aguinaga, and Roger Luje. Detection of behavior patterns through social networks like twitter, using data mining techniques as a method to detect cyberbullying. In *2018 7th International Conference On Software Process Improvement (CIMPS)*, pages 111–118. IEEE, 2018.

[38] Tazeek Bin Abdur Rakib and Lay-Ki Soon. Using the reddit corpus for cyberbully detection. In *Asian Conference on Intelligent Information and Database Systems*, pages 180–189. Springer, 2018.

[39] Michal Ptaszynski, Fumito Masui, Yasutomo Kimura, Rafal Rzepka, and Kenji Araki. Automatic extraction of harmful sentence patterns with application in cyberbullying detection. In *Language and Technology Conference*, pages 349–362. Springer, 2015.

[40] Vinita Nahar, Sanad Al-Maskari, Xue Li, and Chaoyi Pang. Semi-supervised learning for cyberbullying detection in social networks. In *Australasian Database Conference*, pages 160–171. Springer, 2014.

[41] Laura Patricia Del Bosque and Sara Elena Garza. Prediction of aggressive comments in social media: an exploratory study. *IEEE Latin America Transactions*, 14(7):3474–3480, 2016.

[42] Samaneh Nadali, Masrah Azrifah Azmi Murad, Nurfadhlina Mohamad Sharef, Aida Mustapha, and Somayeh Shojaee. A review of cyberbullying detection: An overview. In *2013 13th International Conference on Intellient Systems Design and Applications*, pages 325–330. IEEE, 2013.

[43] Ana Kovačević. Cyberbullying detection using web content mining. In *2014 22nd Telecommunications Forum Telfor (TELFOR)*, pages 939–942. IEEE, 2014.

[44] Elif Varol Altay and Bilal Alatas. Detection of cyberbullying in social networks using machine learning methods. In *2018 International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism (IBIGDELFT)*, pages 87–91. IEEE, 2018.

[45] Selma Ayşe Özel, Esra Saraç, Seyran Akdemir, and Hülya Aksu. Detection of cyberbullying on social media messages in turkish. In *2017 International Conference on Computer Science and Engineering (UBMK)*, pages 366–370. IEEE, 2017.

[46] Thabo Mahlangu, Chunling Tu, and Pius Owolawi. A review of automated detection methods for cyberbullying. In *2018 International Conference on Intelligent and Innovative Computing Applications (ICONIC)*, pages 1–5. IEEE, 2018.

[47] Vikas S Chavan and SS Shylaja. Machine learning approach for detection of cyber-aggressive comments by peers on social media network. In *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 2354–2358. IEEE, 2015.

[48] Walisa Romsaiyud. Expectation-maximization algorithm for topic modeling on big data streams. In *2016 IEEE 7th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, pages 1–7. IEEE, 2016.

[49] Hitesh Kumar Sharma, K Kshitiz, et al. Nlp and machine learning techniques for detecting insulting comments on social networking platforms. In *2018 International Conference on Advances in Computing and Communication Engineering (ICACCE)*, pages 265–272. IEEE, 2018.

[50] Eren Çürük, Çiğdem Acı, and Esra Saraç Eşsiz. Performance analysis of artificial neural network based classfiers for cyberbulling detection. In *2018 3rd International Conference on Computer Science and Engineering (UBMK)*, pages 1–5. IEEE, 2018.

[51] Rekha Sugandhi, Anurag Pande, Siddhant Chawla, Abhishek Agrawal, and Husen Bhagat. Methods for detection of cyberbullying: A survey. In *2015 15th International Conference on Intelligent Systems Design and Applications (ISDA)*, pages 173–177. IEEE, 2015.

[52] Seok-Jun Bu and Sung-Bae Cho. A hybrid deep learning system of cnn and lrcn to detect cyberbullying from sns comments. In *International Conference on Hybrid Artificial Intelligence Systems*, pages 561–572. Springer, 2018.

[53] Mohd Fazil and Muhammad Abulaish. A hybrid approach for detecting automated spammers in twitter. *IEEE Transactions on Information Forensics and Security*, 13(11):2707–2719, 2018.

[54] Eren Çürük, Çiğdem Acı, and Esra Saraç Eşsiz. The effects of attribute selection in artificial neural network based classifiers on cyberbullying detection. In *2018 3rd International Conference on Computer Science and Engineering (UBMK)*, pages 6–11. IEEE, 2018.

[55] Miftah Andriansyah, Ali Akbar, Afina Ahwan, Nico Ariesto Gilani, Ardiono Roma Nugraha, Rizki Nofita Sari, and Remi Senjaya. Cyberbullying comment classification on indonesian selebgram using support vector machine method. In *2017 Second International Conference on Informatics and Computing (ICIC)*, pages 1–5. IEEE, 2017.

[56] Kim D Gorro, Mary Jane G Sabellano, Ken Gorro, Christian Maderazo, and Kris Capao. Classification of cyberbullying in facebook using selenium and svm. In *2018 3rd International Conference on Computer and Communication Systems (ICCCS)*, pages 183–186. IEEE, 2018.

[57] Mengfan Yao, Charalampos Chelmis, and Daphney-Stavroula Zois. Cyberbullying detection on instagram with optimal online feature selection. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 401–408. IEEE, 2018.

[58] Peter Bourgonje, Julian Moreno-Schneider, Ankit Srivastava, and Georg Rehm. Automatic classification of abusive language and personal attacks in various forms of online communication. In *International Conference of the German Society for Computational Linguistics and Language Technology*, pages 180–191. Springer, Cham, 2017.

[59] Adewale Obadimu, Esther Mead, Muhammad Nihal Hussain, and Nitin Agarwal. Identifying toxicity within youtube video comment. In *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*, pages 214–223. Springer, 2019.

[60] Yasin N Silva, Christopher Rich, and Deborah Hall. Bullyblocker: Towards the identification of cyberbullying in social networking sites. In *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 1377–1379. IEEE, 2016.

[61] Djedjiga Mouheb, Rutana Ismail, Shaheen Al Qaraghuli, Zaher Al Aghbari, and Ibrahim Kamel. Detection of offensive messages in arabic social media communications. In

2018 International Conference on Innovations in Information Technology (IIT), pages 24–29. IEEE, 2018.

[62] Trisna Febriana and Arif Budiarto. Twitter dataset for hate speech and cyberbullying detection in indonesian language. In *2019 International Conference on Information Management and Technology (ICIMTech)*, volume 1, pages 379–382. IEEE, 2019.

[63] Wan Noor Hamiza Wan Ali, Masnizah Mohd, and Fariza Fauzi. Cyberbullying detection: An overview. In *2018 Cyber Resilience Conference (CRC)*, pages 1–3. IEEE, 2018.

[64] Francois Mouton, Louise Leenen, and H. S. Venter. Social engineering attack examples, templates and scenarios. *Computers and Security*, 59, 2016.

[65] Marcus Nohlberg and Stewart Kowalski. The cycle of deception: a model of social engineering attacks, defenses and victims. In *Second International Symposium on Human Aspects of Information Security and Assurance (HAISA 2008), Plymouth, UK, 8-9 July 2008*, pages 1–11. University of Plymouth, 2008.

[66] Taimur Bakhshi. Social engineering: revisiting end-user awareness and susceptibility to classic attack vectors. In *Emerging Technologies (ICET), 2017 13th International Conference on*, pages 1–6. IEEE, 2017.

[67] Kevin D Mitnick, William L Simon, and S Wozniak. The art of deception: Controlling the human element of security. 2002. *Paperback ISBN 0-471-23712-4*, 2006.

[68] Patricio Zambrano, Jenny Torres, and Pamela Flores. How does grooming fit into social engineering? In *Advances in Computer Communication and Computational Sciences*, pages 629–639. Springer, 2019.

[69] Shirley Radack. Managing information security risk: Organization, mission and information system view. Technical report, National Institute of Standards and Technology, 2011.

[70] Richard Kissel. *Glossary of key information security terms*. Diane Publishing, 2011.

[71] Ping Chen, Lieven Desmet, and Christophe Huygens. A study on advanced persistent threats. In *IFIP International Conference on Communications and Multimedia Security*, pages 63–72. Springer, 2014.

[72] Brahim ID Messaoud, Karim Guennoun, Mohamed Wahbi, and Mohamed Sadik. Advanced persistent threat: New analysis driven by life cycle phases and their challenges.

In *2016 International Conference on Advanced Communication Systems and Information Security (ACOSIS)*, pages 1–6. IEEE, 2016.

[73] Adel Alshamrani, Sowmya Myneni, Ankur Chowdhary, and Dijiang Huang. A survey on advanced persistent threats: Techniques, solutions, challenges, and research opportunities. *IEEE Communications Surveys & Tutorials*, 2019.

[74] Lin Shenwen, Li Yingbo, and Du Xiongjie. Study and research of apt detection technology based on big data processing architecture. In *2015 IEEE 5th International Conference on Electronics Information and Emergency Communication*, pages 313–316. IEEE, 2015.

[75] Martin Ussath, David Jaeger, Feng Cheng, and Christoph Meinel. Advanced persistent threats: Behind the scenes. In *2016 Annual Conference on Information Science and Systems (CISS)*, pages 181–186. IEEE, 2016.

[76] Eric M Hutchins, Michael J Cloppert, and Rohan M Amin. Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains. *Leading Issues in Information Warfare & Security Research*, 1(1):80, 2011.

[77] Pei-Ju Lee, Ya-Han Hu, Kuanchin Chen, J Michael Tarn, and Lien-En Cheng. Cyberbullying detection on social network services. In *PACIS*, page 61, 2018.

[78] Wei Wang, Payam Mamaani Barnaghi, and Andrzej Bargiela. Probabilistic topic models for learning terminological ontologies. *IEEE Transactions on Knowledge and Data Engineering*, 22(7):1028–1040, 2009.

[79] Ali Daud, Juanzi Li, Lizhu Zhou, and Faqir Muhammad. Knowledge discovery through directed probabilistic topic models: a survey. *Frontiers of computer science in China*, 4(2):280–301, 2010.

[80] Yue Lu, Qiaozhu Mei, and ChengXiang Zhai. Investigating task performance of probabilistic topic models: an empirical study of plsa and lda. *Information Retrieval*, 14(2):178–203, 2011.

[81] Rubayyi Alghamdi and Khalid Alfalqi. A survey of topic modeling in text mining. *Int. J. Adv. Comput. Sci. Appl.(IJACSA)*, 6(1), 2015.

[82] Ethan Fast, Binbin Chen, and Michael S Bernstein. Empath: Understanding topic signals in large-scale text. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 4647–4657, 2016.

[83] Francois Mouton, Mercia M Malan, Louise Leenen, and Hein S Venter. Social engineering attack framework. In *Information Security for South Africa (ISSA), 2014*, pages 1–9. IEEE, 2014.

[84] Aarti Sharma, Rahul Sharma, Vivek Kr Sharma, and Vishal Shrivatava. Application of data mining–a survey paper. *International Journal of Computer Science and Information Technologies*, 5(2):2023–2025, 2014.

[85] Pradnya P Sondwale. Overview of predictive and descriptive data mining techniques. *International Journal of Advanced Research in Computer Science and Software Engineering*, 5(4):262–265, 2015.

6 ON THE MODELING OF CYBER-ATTACKS ASSOCIATED WITH SOCIAL ENGINEERING: A PARENTAL CONTROL PROTOTYPE

Patricio Zambrano[1*], Jenny Torres[1], Luis Tello-Oquendo[2,3], Ángel Yánez[1], Luis Velásquez[1]

[1]Department of Informatics and Computer Science, Escuela Politécnica Nacional, Quito 170525, Ecuador

[2]College of Engineering, Universidad Nacional de Chimborazo, Riobamba 060108, Ecuador

[3]Facultad de Ingeniería en Electricidad y Computación, Escuela Superior Politécnica del Litoral, Guayaquil, Ecuador

6.1 ABSTRACT

Nowadays, the psychological techniques used to harass, intimidate, threaten, steal information are more common due to free access to technological resources and the digitization of communications. However, studies related to cybersecurity concerning social engineering techniques are still limited. Several factors such as access to specific databases on cyber-attacks, the unification of scientific criteria that evaluate the nature of the problem, or the absence of accurate proposals that prevent and mitigate this problem could motivate researchers'lack of interest in information security to generate meaningful contributions. This research presents a cyber-attack modeling process defining its stages through topic modeling. This process is applied to model grooming and bullying cyber-attacks, in which the use of psychological manipulation techniques by the attackers was evidenced. The proposal allowed the detection of the attackers'communicational intentions with high accuracy. Additionally, a functional parental control prototype that supports the proposed modeling process is presented.

KEY WORDS: APT, Bullying, Cyber-attack, Grooming, Pattern Behavior, Social Enginee-

ring, Topic Model

6.2 INTRODUCTION

At present, online harassment is a concept that has demonstrated its presence in various communication channels, although it has not been approached globally. Attacks such as grooming, bullying, gender violence, bank fraud, among others, share the same pattern: psychological manipulation to obtain benefits such as personal or financial satisfaction. This entity is analyzed as part of the cybersecurity field from the information security perspective. The openness and availability of digital communications allow people to express their opinions freely. However, sometimes this content includes traits of harassment of all kinds, such as rejection and hatred. This emotional charge reflected in digital textual content requires new proposals supported by computerized mechanisms for its study. It should be noted that the increasing incidence of these cyber-attacks has psychological and economic repercussions. For the first case, a high risk of suicidal behavior has been evidenced in the absence of preventive mechanisms that allow the first signs of an attack to be observed. It is worth mentioning that innovative technological solutions are not enough to combat this type of attack. Collective and collaborative work between the family, police, and judicial spheres is necessary to generate more effective defenses against these cyber attacks.

In today's environment, the virtual communications usability trend is increasingly evident. A fundamental challenge for researchers is to propose mechanisms that allow managing the knowledge inherent in virtual communications. For this purpose, natural language processing (NLP) techniques are necessary. One of the most widely used NLP techniques is topic detection and modeling. Several algorithm proposals, such as pLSA, LDA, HDP, TDM, NMF, among many others, determine connecting aspects (topics) between the words that make up digital texts. There is free access and licensed libraries that have been previously analyzed and tested in the scientific field to implement these algorithms. In the classification of digital information, those texts are considered "short texts"; these come from comments, blogs, and chats. Short texts do not have a defined structure; however, they allow defining common patterns in large quantities.

This research proposes a modeling process of an attack related to psychological violence. Large amounts of texts related to online harassment were used for this purpose. They went through debugging and modeling processes to find intrinsic patterns that allow modeling

cyber-attacks such as grooming and bullying from information security. The results with a high degree of precision supported the feasibility of applying the process to new attacks related to the study of online harassment in short texts (chats).

The main contributions of this study are summarized as follows.

- ❖ A framework to model cyber attacks related to bullying and grooming;
- ❖ Propose current techniques for obtaining and processing data related to instant messaging;
- ❖ Propose alternative models of topics in short word processing;
- ❖ Point out functional tools that allow establishing lexical meanings to the topics obtained in the modeling process;
- ❖ Show the architecture and operation of a parental control prototype based on the proposed framework.

The remainder of this paper is organized as follows. Section 6.3 presents background information that allows clarifying the research problem. Section 6.4 describes the proposed framework and the process carried out to model grooming and bullying cyberattacks, respectively; we describe the techniques, tools, models and other aspects that were used in each phase of the process. Section 6.5 elaborates on a functional parental control prototype implemented based on the proposed framework. Section 6.6 answers the research questions posed. Finally, Section 6.7 draws conclusions and presents future work.

6.3 PROBLEM UNDERSTANDING

6.3.1 Background

6.3.1.1 Cyber-attacks vs. Assets

Today, one of the organizations'most valuable assets is information, and various strategies or controls are used to prevent it from being affected by unwanted attacks. However, this concept is exclusively aligned with information security and not with cybersecurity. The authors

in [1] explore the different definitions of information security and cybersecurity. They describe that, although the concept of cybersecurity is aligned with that of information security, their coverage is different. Information security is the protection of information considered an asset that is susceptible to potential attacks. On the other hand, cybersecurity covers the protection of cyberspace itself and each of the elements that generate information (assets).

Currently, several cyber attacks guide their attention to the harassment of people; cases such as cyberbullying and grooming are present in this spectrum of unconventional attacks. In various studies [1], [2], [3], the increasing trend of these attacks is demonstrated. In their eagerness to compromise their victims emotional and psychological stability, the aggressors use intimidation and harassment strategies so that their victims feel shame and normalize violence. Physical damage is achieved in the most advanced cases, generating severe, negative, and negative impacts irreversible in victims. All these processes are carried out with technological tools. Scientists and governments are developing proposals that progressively address and mitigate these cybersecurity attacks, given the rise of cyberbullying. It should be noted that being a victim of cyberbullying in cyberspace does not establish a loss of confidentiality, integrity, or availability of a type of tangible information. Instead, the target of these attacks is the victims and their emotional stability. Consequently, people and their physical-emotional stability must be part of the concept of assets within the field of cybersecurity.

6.3.1.2 Psychology vs. Social Engineering

In the area of computer security within Computer Science, there is Social Engineering. This field of study is responsible for studying and establishing the techniques or practices of psychological manipulation that attackers use to obtain confidential information from computer resources or people. The increase in attacks concerning tactics related to social engineering in cyberspace has led to research linking the fields of psychology and technology. Such is the case, in [4, 5] the authors generalize the basic principles of social engineering based on the behavior and susceptibility of victims. In these investigations, the authors adapted knowledge of experimental psychology to identify factors that increase the probability of success of a social engineer against a human victim.

6.3.1.3 Technology vs. Psychology

Solving social demands related to cognitive processes and interpersonal relationships is one of the objectives of scientists in the field of psychology. Furthermore, the application of information technologies has made it possible to generate new knowledge in this area. For this reason, technology has become a fundamental tool that psychology uses for the benefit of people through different techniques of analysis, evaluation, and modification of human behavior [6, 7].

In recent years, social networks show the growing interrelation of information technologies and the psychological effect on people. This trend allows specialists in the psychological field to study human behavior and mental health concerning the use of technologies. As a case study in [8], the authors compiled research related to the use of social networks and the interaction between individuals in them; however, they point out that the data is still limited. Therefore, this compendium of research analyzes and identifies latent meta-groups of online communities with and without mental health-related conditions where aspects such as depression and autism are part of the analysis of results.

With these antecedents, it is observed that psychology makes use of information technologies to develop new knowledge. However, the field of psychology has also made it possible to strengthen concepts, criteria, and theories that have been born from computer science. Within this field, there is much digital information related to social behavior. This information requires formal processes that allow the establishment of common patterns aligned with the field of psychology. For this reason, for several years, various proposals have been analyzed to standardize the behavior of online bullies (grooming), supporting their results with psychological concepts [2, 3].

6.3.1.4 Subjectivity in the scientific results

In studies related to online bullying, the psychological component is observed. This component allows the establishment of the different phases that are part of the operation of the attack; however, it is believed that this human component has not allowed establishing a standard procedure that allows modeling other attacks with similar characteristics due to its high level of subjectivity. In [9] this aspect is already evidenced as an obstacle to the research validation since the peer review is based on objective and not subjective data. On

the other hand, it is stated that there are research areas where the lack of availability of adequate research tools and the difficulties associated with quantifying subjective data do not allow establishing conclusive data [10]. In our case study, it was established that there is a set of tools based on statistical algorithms that allow reducing this subjectivity [2, 3]. With the use of topic modeling, relevant information can be extracted through categories or topics from large volumes of digital information, which in our case are instant messaging chats.

6.3.1.5 Advanced persistent threat modeling - APT

In the technical description, [2], advanced persistent attacks have a wide range of proposals. Models such as Lockheed, LogRhythm, Mandiant, Dell Secureworks, SDAPT, BSI, and Lancaster vary their phases. This difference in criteria is due to the lack of agreements or standardization to model APT attacks. These models base their phases on experiential aspects during the analysis of the various attacks. On the other hand, from the social engineering perspective, alternative models to APTs have been proposed, such as the case of Kevin Mitnick's 4 phases (information gathering, relationship development, relationship exploitation, and execution to achieve the objective) [11], and Mounton (attack formulation, information gathering, preparation, relationship development, relationship exploitation and debrief) [12]. To standardize the criteria for modeling attacks related to social engineering related to APT concepts, Zambrano et al. [2], in their proposals, relates the phases of the models with the lexical categories determined in each evaluated attack, grooming, and bullying cases.

6.3.2 Review of Topic Modeling Applications

Topic modeling is an unsupervised machine learning technique, which is part of the concept of artificial intelligence and specifically natural language processing (NLP). This has the particularity of analyzing large sets of documents, detecting common patterns between the words that make them up, and grouping them into topics representing them. Factors such as coherence and perplexity allow us to study these groupings of words. In the field of text analysis, inspired by artificial intelligence, there is a wide range of methods or algorithms that process information, such as topic analysis.

Currently, several articles make comparisons and describe the applications of different models that are part of the modeling of topics, see Table 6.1. These surveys analyze the current

usage trends and structural aspects of the models [13]. In [14] the behavior of several models with specific data is analyzed, called by the authors “short texts.” The results discuss the reliability of grouping this type of poorly structured information in the resulting topics. On the other hand, the authors in [15] classify the existing topic models into two categories: topic models (e.g., LSA - LDA) and topic evolution models, which model topics considering a time factor (e.g., TOT - DTM).

In [16] a variation to LDA is proposed, called Online LDA (OLDA), whose approach provides an efficient means to track topics over time; it is essential to highlight that it has detection characteristics of emerging topics in time. This model was evaluated qualitatively and quantitatively. Another application example is given in [17] where they use the OLDA features to group scientific articles published over time.

It is worth mentioning that in [2], [3] reviews of the literature are shown in greater depth. These are related to each case study. On the other hand, the authors consider that a success factor of previous research was the use of widely tested, updated algorithms that are available for reproducibility in future research.

One of the limitations of the results of the modeling of topics (word bags) is the lack of description of a lexical meaning. These models have an algorithmic mechanism that groups words according to the inherent grammatical aspects of each document. In previous research (Bullying and Grooming), the authors propose using the EMPATH system or the LIWC dictionary to establish meanings or lexical descriptions for each bag of words resulting from the applied model.

6.3.3 Research questions

Formalizing the cyber-attack modeling process with concepts related to traditional information security will allow the scientific community to support future research related to social engineering, topic modeling, and analysis of social patterns. To achieve these goals, we formulated the following research questions:

RQ1: Is the proposed modeling process applicable to other cyber-attacks that evidence social engineering tactics?

RQ2: Which topic model presents better results in the evaluation of short texts?

Table 6.1: Topic Modeling - Literature review

Literature review related to the use of topic modeling in research	***Reference***
Topic models using semantic context to improve document classification	[18], [19], [20], [21], [22], [23], [24], [25], [26]
Topic modeling applied to the study of short texts	[27], [28], [29], [30], [31], [32], [33], [34], [35]
Topic models that identify topics as they emerge over time	[36], [16], [37], [17], [38]
Incorporation of topic features to improve document grouping accuracy	[39]
Multi-part topic model improves information retrieval and document classification performance	[40]
Topic models that automatically detect recurring patterns of expressions	[41]
Modeling of topics using word co-occurrence patterns.	[42]
Word embedding for improved topic modeling	[43], [44], [45] [46], [47], [48]
Incorporation of phrases into topics models to add coherence	[49], [50]
Parallel computing as topic modeling	[51]
Topic modeling with artificial neural networks.	[52]
Topic modeling for short texts using word patterns	[53], [54]
Topic models that obtain contextualized summaries	[55]
Topic models that perform an analysis taking into account human attention	[56]
Topic models that filter and reduce generality and impurities in documents	[57], [58], [59]
Topic model that takes into account entities that are present in documents	[60]
Topic modeling that exploits the sequential order of sentences and the relationship between successive sentences	[61]
Topic model focused on solving data scarcity and improving data security	[62]
Topic modeling with focus on citations and document titles	[63]
Topic models implementing hierarchical topic trees	[64]
Topic model linking short texts to long documents	[65]
Literature review of research related to LDA	***Reference***
Multi-part topic model improves information retrieval and document classification performance	[40]
LDA model applying distances with probability distributions	[66], [67], [68]
LDA model applying weights to words in the sample	[69], [70]
Topic model adding a category distribution parameter to LDA	[71]

RQ3: Can the models determined by the proposed process be implemented in functional systems today?

In the following, we introduce a generalized framework to model cyber-attacks related to social engineering. In Section 6.4.1, we develop the modeling process based on four stages; then, we present the procedural application on two common attacks, namely bullying and grooming in Section 6.4.2.

6.4 FRAMEWORK PROPOSAL TO MODEL CYBER-ATTACKS RELATED TO SOCIAL ENGINEERING

This section introduces a method for modeling an attack involving psychological violence. For this, a large number of texts about internet harassment were used. We used debugging and modeling techniques to discover intrinsic patterns that can be used to model cyber-attacks like grooming and bullying from the perspective of information security. The results showed that the approach may be applied to new attacks connected to the study of online harassment in brief texts (chats) accurately.

6.4.1 Cyber-attack modeling process

The process to model cyber-attacks that involve psychological violence comprises four stages as depicted in Fig. 6.1: Attack selection, Data debugging, Topic model application, and Model description.

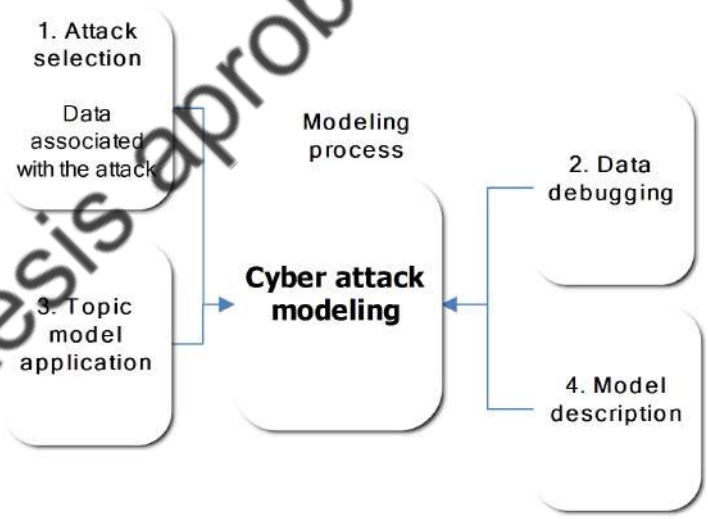


Figure 6.1: Cyber-attack modeling process

1. Attack selection, see Figure 6.2.- This sub-process establishes that: any investigation related to cyber-attacks must go through a process of analysis of the literature. This analysis verifies the existence of previous studies or proposals that propose data download procedures or specific modeling. In this sub-process, the researchers must collect texts, chats, or any information related to instant messaging. These data must

demonstrate the use of psychological manipulation techniques or social pressure by the attackers for the investigation to be viable. The data collected must go through previous debugging processes.

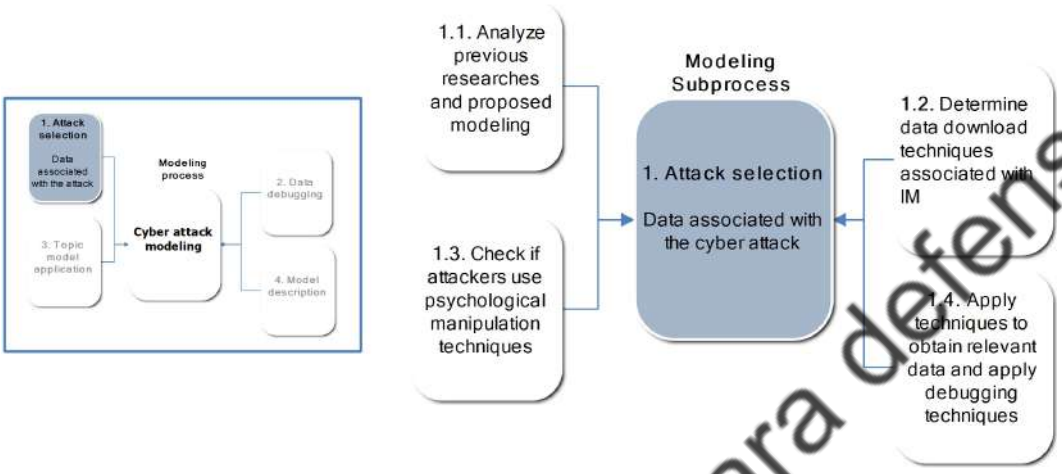


Figure 6.2: Subprocess - Attack selection

2. Data debugging, see Figure 6.3.- In this sub-process, researchers must rely on standardized data mining models. These have recommendations for the treatment and analysis of large information bases. Models such as CRISP-DM (Cross Industry Standard Process for Data Mining) or SEMMA (Sample, Explore, Modify, Model, and Assess) specify the tasks to be carried out in each phase described by the information processing process.

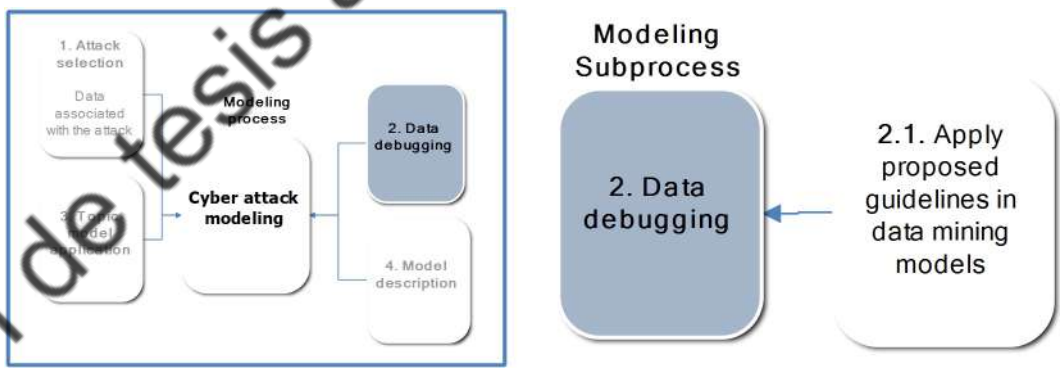


Figure 6.3: Subprocess - Data debugging

3. Topic model application, see Figure 6.4.- In the area of natural language processing, within the Artificial Intelligence (AI) models, various algorithms are described that allow modeling of topics in such a way that unsupervised large volumes of information. Models such as LSI (latent semantic indexing), LDA (Latent Dirichlet Allocation), HDP (Hierarchical Dirichlet Process), and NMF (Non-negative matrix factorization), among

others, are available for use and application in Python and MATLAB. It is worth mentioning that any investigation must go through a rigorous analysis of these models, in contrast to the data, so that the results obtained have a higher level of accuracy and a lower number of word groupings (topics). It is important to take into account that these models have certain peculiarities that associate them. On the one hand, there is the type of data that must be entered into these models. These must previously be refined and lemmatized for modeling. This modeling will be associated with the assignment of a standard number of topics that the researcher must define for each cyber attack. Variables such as perplexity, computational cost, or coherence will determine a specific number of word groupings to refine the final models.

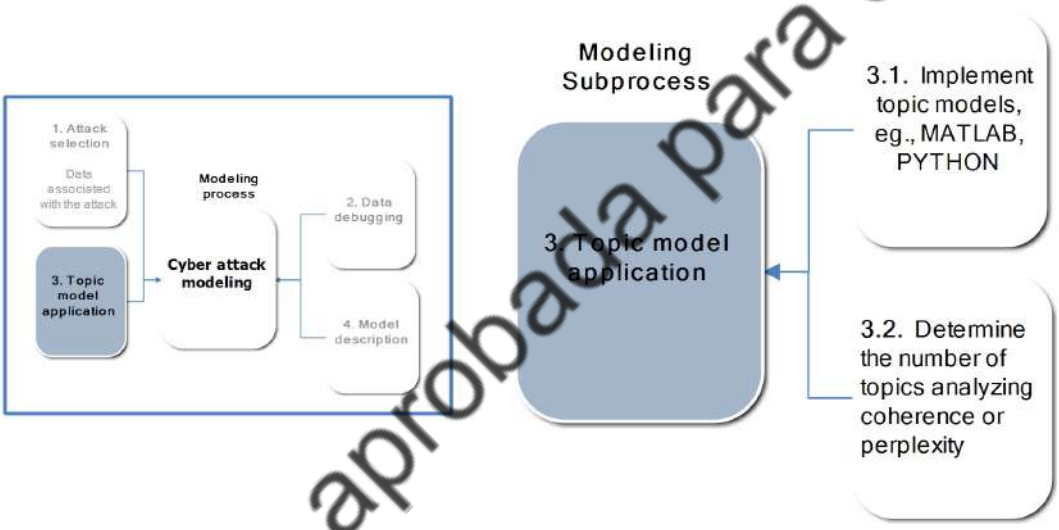


Figure 6.4: Subprocess - Topic model application

4. Model description, see Figure 6.5.- Once the word groupings are obtained, they do not have an explicit meaning. With the use of software tools or manual dictionary-based categorization, lexical categories and communication intentions are defined. This information, in the future, will be related to APT model stages and their concepts.

6.4.2 Application of the modeling process to cyber attacks

Currently, some cyber-attacks make use of psychological manipulation. Attacks such as grooming and bullying have evolved, from a technical point of view, in recent years. On the other hand, it has been observed that the research associated with studying these phenomena has not grown in the same way for their identification and mitigation. The causes that

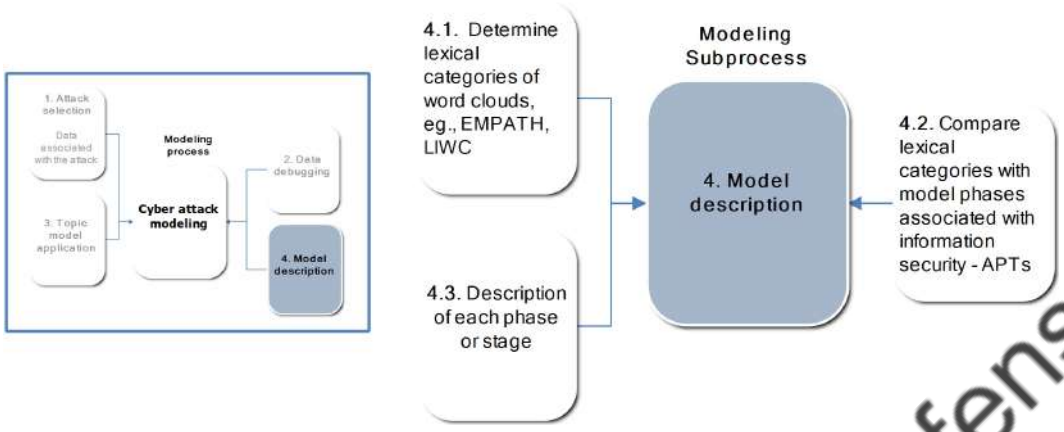


Figure 6.5: Subprocess - Model description

prevent this growth may be due to the confidentiality to publish the results of the investigations, access to data related to the attacks, legal regulations of each country, and even lack of agreements to standardize the modeling to be applied in future proposals. In [3], the authors consider that these aspects can be addressed to increase the proposals in the area. Regarding standardization, it has been determined that cyber-attacks can be modeled from the point of view of information security. Data can be collected from various freely accessible sources without this violating legal regulations or threatening the privacy of minors. Under these premises, the applicability of the proposed modeling process to the attacks of Bullying and Grooming will be justified.

6.4.2.1 Attack selection

Grooming and Bullying cyber-attacks were analyzed and developed from the perspective of Social Engineering, since the use of psychological manipulation techniques by the attackers was evidenced. The detailed description of these attacks can be found in [2, 3]. This section describes the procedural application of the modeling process to the case studies, specific details can be seen in Table 6.2.

6.4.2.2 Data debugging

In this section, the CRISP-DM model is adapted (see Figure 6.8) to Grooming and Bullying case studies, respectively.

Table 6.2: Application of CRISP-DM to Grooming and Bullying

	Grooming	Bullying
Previous analysis of the literature	Yes	Yes
Download mechanisms were taken into account to obtain the data	No	Yes (partially)
Data source	perverted-justice.com	twitter.com hatebase.org pacerteensagainstbullying.org pacerkidsagainstbullying.org
Type of data	Chats - HTML	Chats - HTML Experiences - HTML
Amount of data	128171 chat lines 100 conversations	250000 attacker-related tweets 3035 victim experiences
Download mechanism	Manual downloads Script developed in python	Scrapy, see Figure 6.6 Script developed in python
Analysis of data	Only from attackers (tweets) Only victims (experiences, see Figure 6.7	

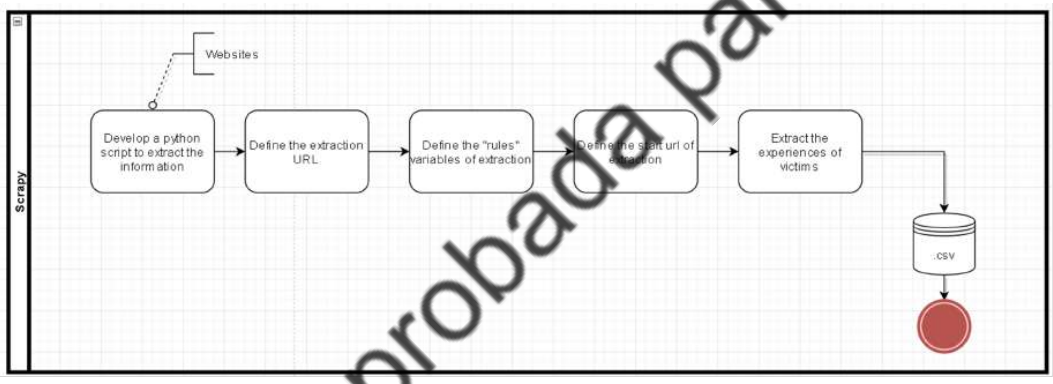


Figure 6.6: Scrapy technique to download short texts (tweets)

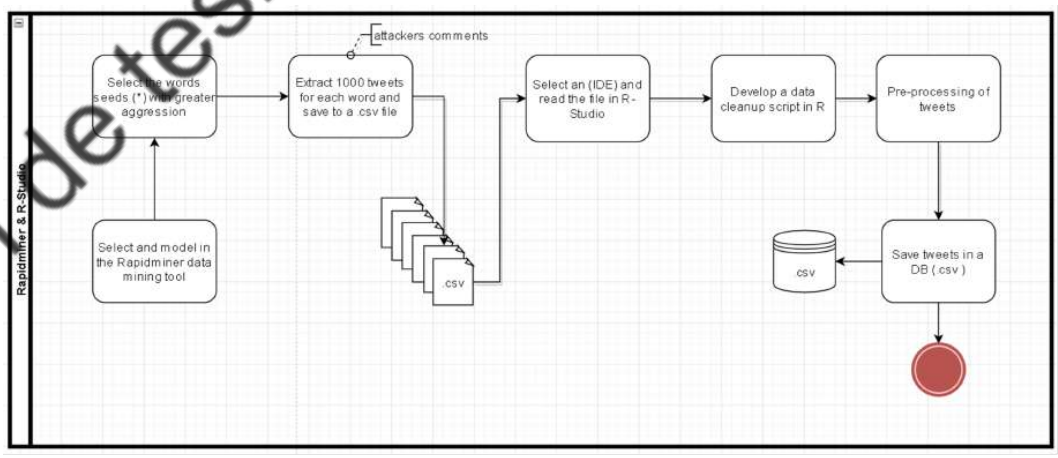


Figure 6.7: Rapidminer and R Studio technique to download experiences of victims

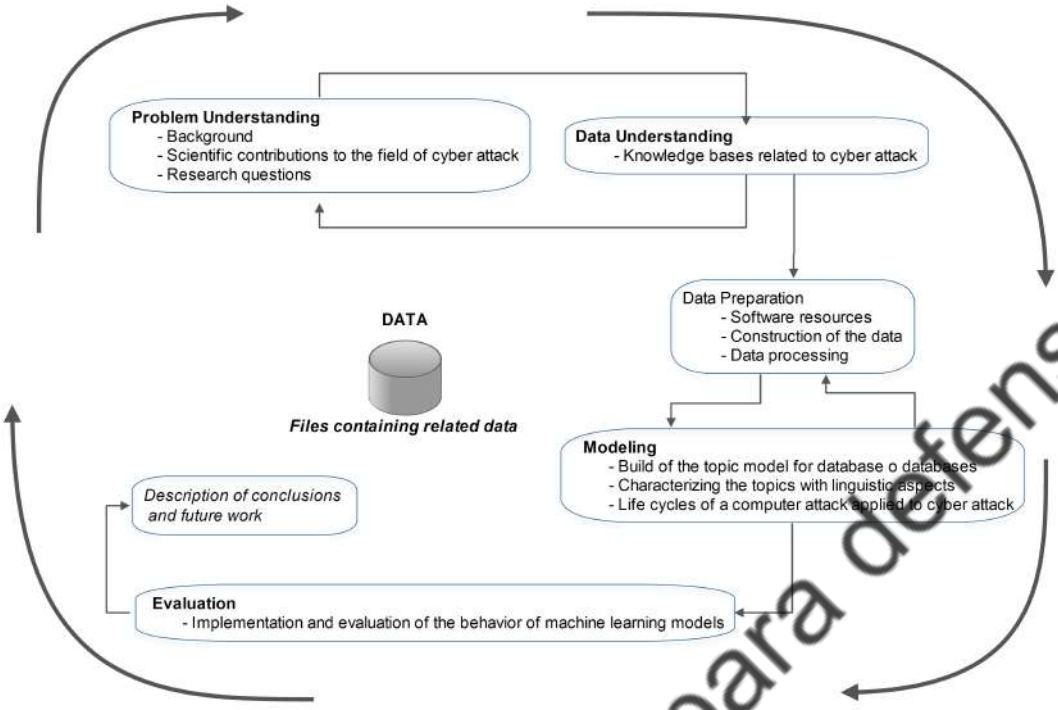


Figure 6.8: CRISP-DM model applied to cyber attacks

6.4.2.3 Topic model application

In the Grooming and Bullying case studies, LDA topic modeling was applied, see Table 6.3. Various studies, in their comparisons, showed that this model provides better results in the analysis of short texts. This supported the application of this model to the cases described [67, 42, 45].

Table 6.3: Topics defined in the analysis tools

	Grooming	Bullying
Type of data	Short Text - Chats	Medium Text - Experiences
Applied model	LDA	LDA
Tools	Matlab Text Analytics	Python Gensim
Aspect used to define the number of topics	Perplexity	Perplexity Coherence
Number of topics defined	6	4

Regarding the scope of the applicability of the topical models, the data behavior with three different models to LDA will be demonstrated below. The implementation of these models was carried out with Python, and the databases analyzed in predecessor investigations were used for this purpose.

1. LSI is a topic model proposed by Deerwester et al. [72]. It is part of a set of natural language processing techniques, particularly distributional semantics. LSI is an indexing and retrieval method that uses a mathematical technique called SVD singular value decomposition, making it possible to determine patterns in a set of words in a collection of texts. SVD supports its operation in classical techniques of second-order eigen vectors and aims to analyze large volumes of information and look for components that are not correlated. LSI has a probabilistic variant developed by Thomas Hofmann known as probabilistic latent semantic analysis, PLSI, or PLSA. This model started the development of LDA, which is the Bayesian version of PLSI.

2. LDA model developed by Blei et al. [73]. It classifies texts from various documents into topics. LDA models its topics with Dirichlet distributions. In this process, word-topic data arrays are established using the Gibbs sampling algorithm of Dirichlet distributions. Each sample will give the probability of each word by topic.

3. NMF is a model proposed by Lee et al. [74]. This model is made up of 2 methods, one performs the dimension reduction functions, and the other performs a factor analysis. The factoring process allows a weighting based on the semantics between the words. Applying an optimization process, the model monitors that within its data, there are no negative values. Finally, a matrix of weighted terms is obtained and grouped in their respective topics.

4. HDP is a mixed model that performs an unsupervised analysis of pooled data. This model was proposed by Teh [75]. Unlike LDA, HDP infers the number of subjects from the data. With the use of variational Bayes coordinate ascending algorithms, HDP manages to optimize the processed data stochastically. Unlike its predecessor LDA, HDP is unlimited in defining the number of topics and learns from your data without the need to pre-specify the number of topics.

Figure 6.9 schematizes the behavior of the data (related to grooming) applied to the proposed models LSA, LDA, NMF, and HDP. The models applied the six topics, defined, and justified by the predecessor investigations. The results show that the LSI and HDP models generate greater dispersion between the coherence values obtained in each topic. Discarding these two models from the analysis, the LDA model and its HDP variant concentrate the processed data better.

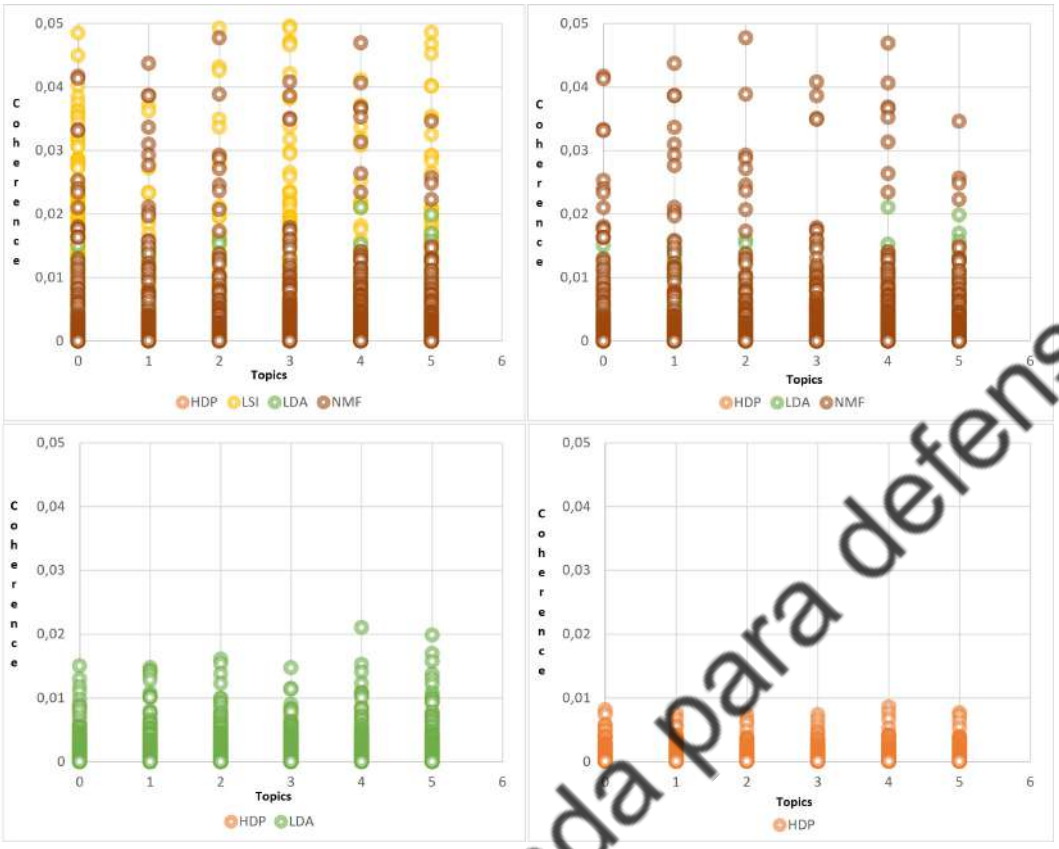


Figure 6.9: Graphical comparison of the behavior of the coherence values in 4 topic models - Grooming

Regarding Bullying, see Figure 6.10, the behavior of the models is similar to the case of Grooming with the difference that it has four topics. Similarly, the LSI and NMF models are discarded since the coherence values are very dispersed concerning LDA results and its predecessor HDP. The HDP model, when applied in the 2 case studies, offers better results in the classification of words by topic. It is worth mentioning that HPD is an improved model to the LDA. With this premise, it is concluded that the LDA model or its variants (HDP) are adequately adapted to short texts, as shown by the research described above.

6.4.2.4 Model description

With the results obtained when applying LDA, we proceeded to define the lexical meaning of each group, considering the communicational intention of the attacker. Table 6.4, concerning Grooming, indicates that the process to determine the lexical meaning of each topic had several stages.

In the process of describing each topic lexically, the categories proposed in LIWC were used

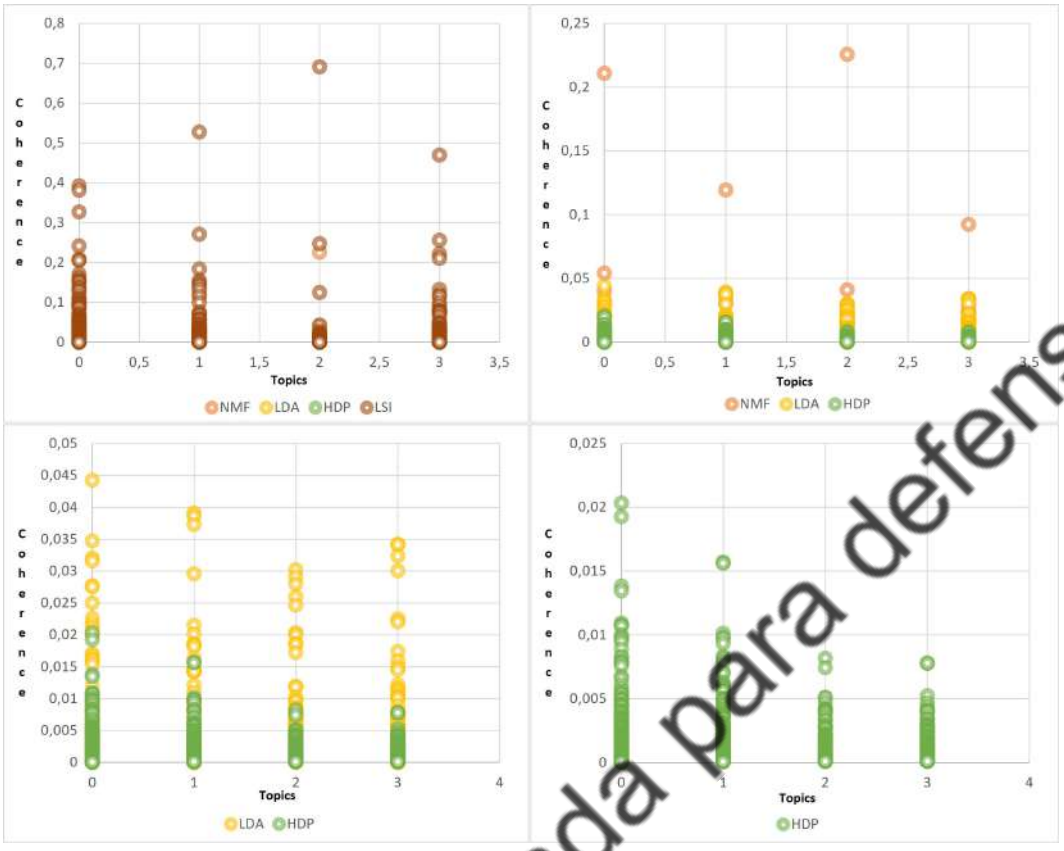


Figure 6.10: Graphical comparison of the behavior of the coherence values in 4 topic models - Cyberbullying

Table 6.4: Linguistic tools and security models categorization

	Grooming	*Cyberbullying*
Number of topics	6	4
Lexical Categories	LIWC Manual classification	Empath Automatic classification
Analysis of communicational intentions	Manual	Manual
Selected topics	Gethering informatión (SDAPT) Gaining Access (SDAPT) Lateral Movement (LogRhythm) Escalating Privileges (Mandiant) Execution (Mitnick) Debrief (Mounton)	Development of relationship (Mitnick) Preparing/Distracting Attack (BSI) Exploit the Relationship (Mounton) Debrief (Mounton)

to describe the communicational intentions of Grooming. It should be noted that this tool was developed for the cognitive and emotional evaluation of texts through a series of psychological and structural categories. This program, used in the field of psychiatry, analyzes texts, word by word, in a classification of different linguistic variables, which include standard language categories (articles, prepositions, pronouns, among others), psychological processes (categories of positive emotions and negative, cognitive variables), words related

to space-temporal relativity, verb tenses, and traditional dimensions of content organized hierarchically [76].

On the other hand, EMPATH is a tool that can generate and validate new lexical categories on demand from a small set of seed terms. EMPATH sketches connotations between words and phrases using deep learning and neural integration in more than 1.8 billion words of modern fiction. EMPATH can analyze text in 200 pre-validated and built-in categories. In [77] they demonstrate that the categories validated by humans and based on data from EMPATH are highly correlated ($r = 0.906$) with similar categories from LIWC.

With the use of the tools described, we proceeded to define a lexical characteristic for each word grouping and compare these in terms of communicational intentions with the stages of the models proposed in information security as follows:

1. Manual comparison of the words of each topic with the dictionary of the LIWC tool.
2. The dictionary classifies certain words into categories, and these were selected based on the data.
3. Once the categories that most contextualized the type of grouped data had been selected, we proceeded to analyze the communicational intentions that defined these categories.
4. With this analysis, the communicational intentions of each topic were compared with the proposed phases of the different models applied to APTs.

Unlike the Grooming study, the Cyberbullying data was analyzed by the WEB EMPATH tool, which automatically classified the words into lexical categories. With these categories, item 3 of the Grooming study was continued to determine the phases from the perspective of information security.

6.5 IMPLEMENTATION OF A PARENTAL CONTROL PROTOTYPE

We present a general architecture of a parental control prototype devised to solve different social problems based on cyberbullying. As this type of attack increases in the number of victims and the complexity of the analysis of the behavioral pattern of the attackers, it is necessary to establish viable technological alternatives to minimize the risk of attack.

In general, the proposed architecture is developed by decomposing the system into components that gather specific aspects of it as a result of an abstraction process and that, when organized in a certain way, constitute the basis of the proposed prototype. In this way, the architecture of the parental control prototype can be seen as the system's structure based on the definition of the components and their interactions; they are described in blocks as depicted in Fig. 6.12.

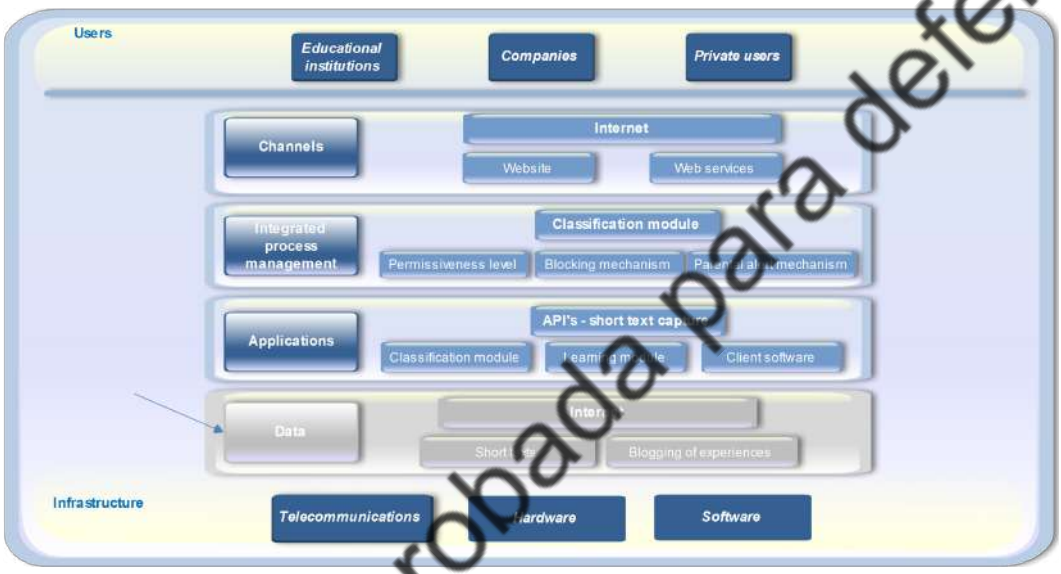


Figure 6.11: Architecture of the parental control prototype

The proposed system is a technological solution based on the proposed modeling process (see Section 6.4), where the number of phases of the attacks is determined through the modeling of topics, suggesting communicational intentions to them. These new structured and tagged data will be part of the system's learning to classify future texts and determine if they are part of attacks related to cyberbullying. It is worth mentioning that the proposed architecture is scalable since it supports the growth of the databases, change of learning model, communication with mobile applications, and allows the entry and management from different perspectives. The implementation of the system contemplates two architectures: WEB under the figure REST and Client-server. Note that this research's main contribution focuses on formalizing the process of selecting, grouping, and labeling the data (Database and Learning Modules) so that they can be processed by any machine learning mechanism based on the attackers'behavioral patterns evidenced.

6.5.1 Operational component

The blocks corresponding to the main components of the application in contrast to the architecture of the prototype are illustrated in Fig. 6.12. The potential users of interest, necessary infrastructure, and the set of tools that support the processes of the proposed prototype are described below.

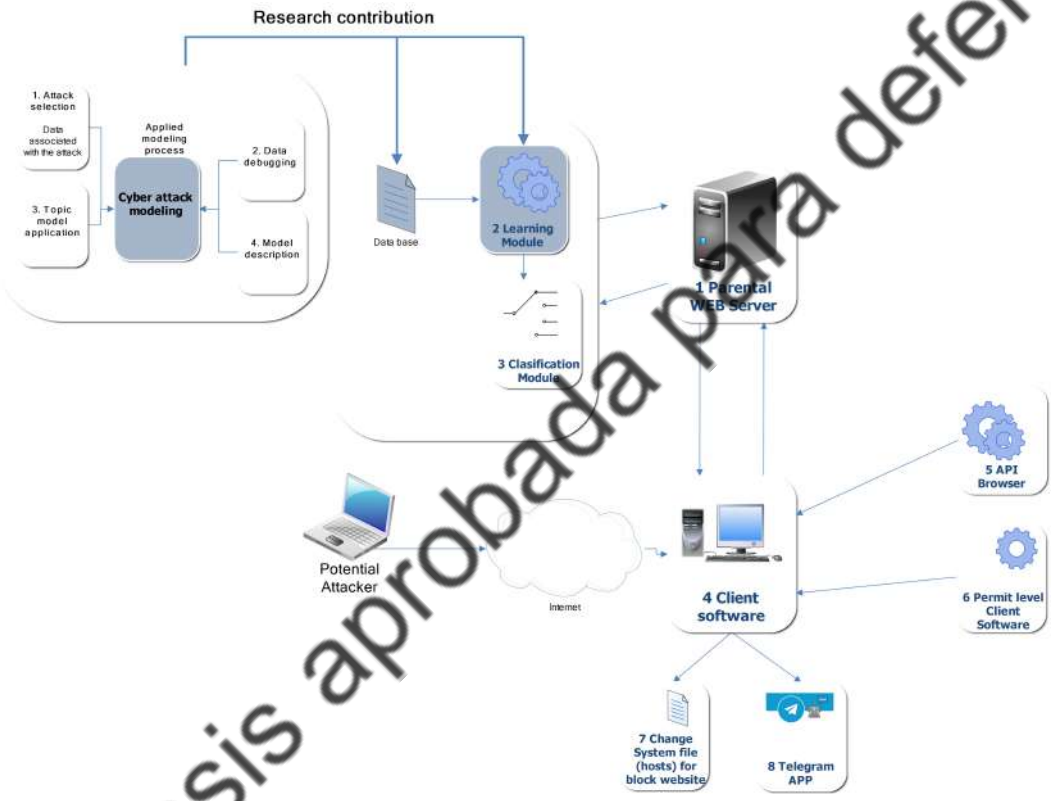


Figure 6.12: Main components of the prototype

6.5.1.1 Data

After acquiring chats (short text data) or experiences of victims who have suffered some type of cyberbullying, the cyberattack modeling process is applied (main contribution of the research). From this processing, grouped data are obtained and labeled according to the communicational intention that defines them.

6.5.1.2 Application

Once the data has been cleaned, classified, and labeled, an automatic learning algorithm is implemented. In the particular case of this prototype, the linear classification algorithm was implemented, which in previous studies [2, 3] obtained better results. With this, the system will classify and label future data coming from the client software, which has a mechanism for extracting chats from the Facebook social network in real-time through an API implemented in the Google Chrome browser.

6.5.1.3 Integrated process management

The API will supervise the sending of the chats in real-time to the parent server to automatically classify them in the programmed topics using labels, see Fig. 6.12. The client module will receive the tagged text and compare the number of the label with the level of system permissiveness (topic configured by the user); see Figures 6.13 and 6.14. If this number exceeds the allowed topic, the system will immediately block communication with the origin website (Facebook).

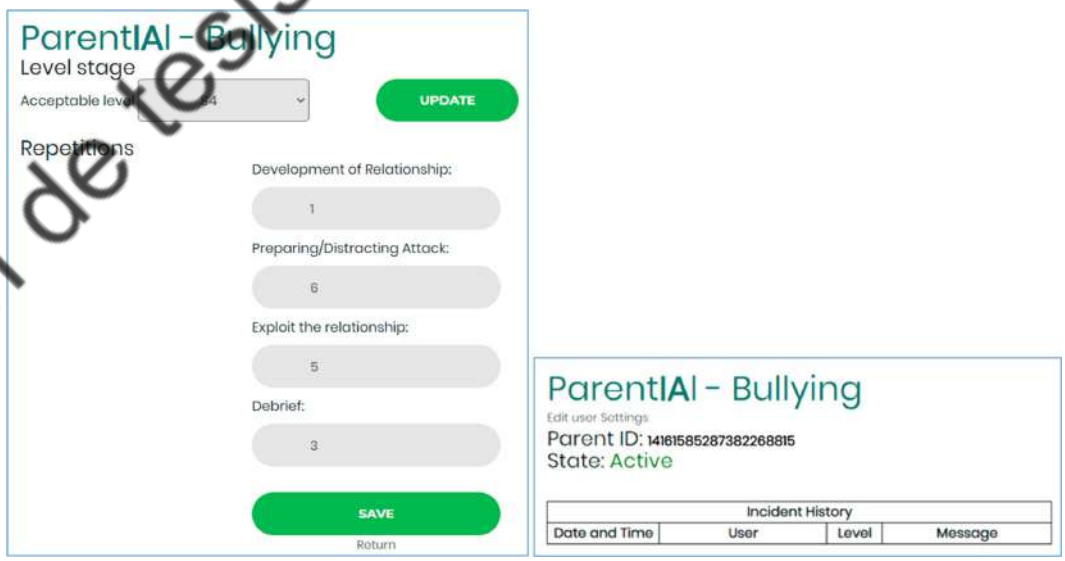


Figure 6.13: Modeling process: system permissiveness

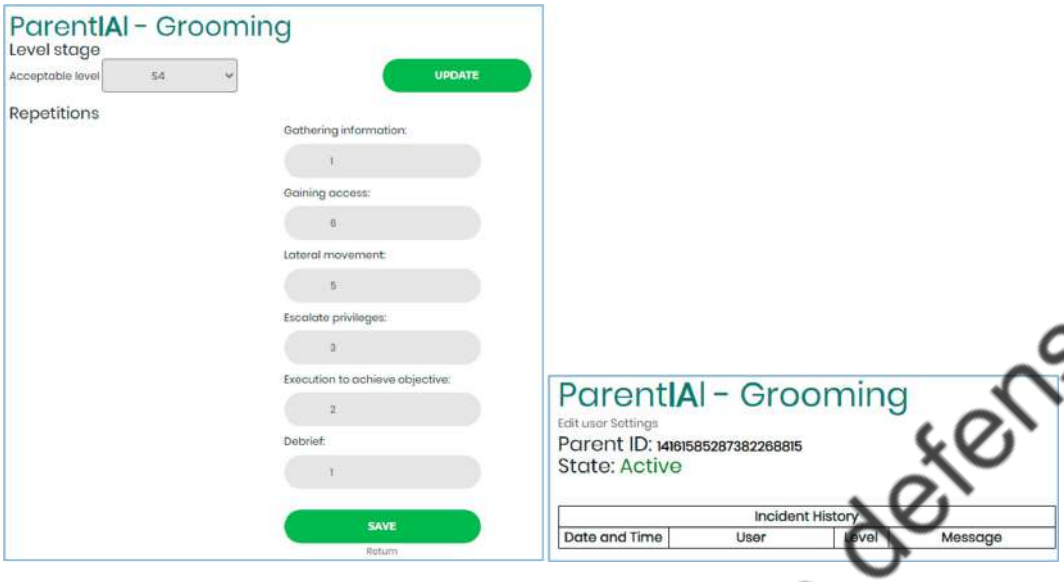


Figure 6.14: Modeling process: system permissiveness

6.5.1.4 Channels

After a comparative analysis of labels, the texts considered malicious would be sent, over the Internet, via telegram text message (Web services) to those in charge of supervising the victims, see Fig. 6.15, notifying the level of aggressiveness in where the attack is located. In addition, this message will contain the attacker's username, time, and date of the event.

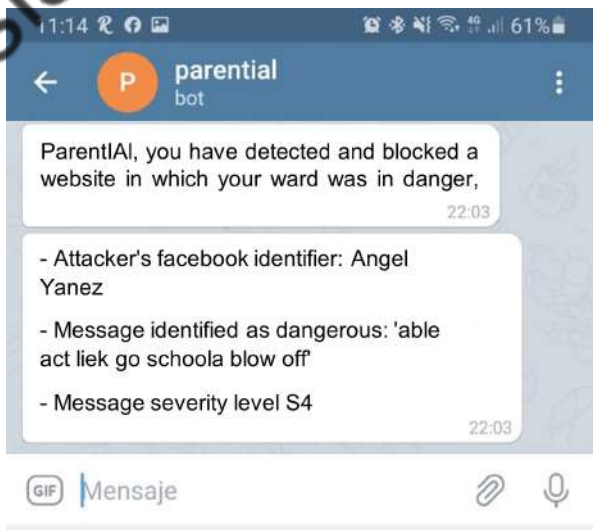


Figure 6.15: Notification

6.5.2 Results'summary

Table 8.1 summarizes the results of the theoretical component and the application part of the previous investigations, Grooming and Cyberbullying. As can be seen, the data processed for Grooming, in the first instance, was based on short texts from chat lines. In the case of Cyberbullying, these data were extended to more enriched texts (experience blogs). These data were preprocessed by applying the CRISP-DM methodology to different topic modeling. In the case of Grooming, these results could be contrasted with theoretical results. These results, number of topics, were obtained based on the analysis of perplexity and computational cost, defining a compromise solution that determined 6 (Grooming) and 4 (Cyberbullying) groupings of words that later formed part of phases of the life cycles of each attack. The investigations consolidated their results with practical aspects applying machine learning techniques. The final labeled data was sectioned. One portion was part of learning and the other for testing, thus determining the accuracy of machine learning techniques against unknown data.

Table 6.5: Summary of previous research

Theoretical Component						Research Application		
Research	Dataset		Average of topics proposed in the literature	Compromise solution Topics obtained	Applied models to get topics	AI techniques for learning	Dataset sectioning to learning & Tests	Model accuracy result
Grooming	100 conversations 128171 chat lines		6	6	LDA	LinearSVC CNN LSTM	90 % - 10 %	97,61 % 96,11 % 95,91 %
Bullying	250,000 attacker related tweets	3035 victim experiences	-	4	PLSA LDA	LinearSVC LSTM	70 % - 30 %	94,74 % 94,41 %

Table 8.3 describes how the prototype was developed with a theoretical and practical component. The theoretical component, unlike previous research, has an evaluation of different models related to topic modeling to contrast the results of previous research. With this background and the results obtained in the implementation of Grooming and Bullying, a prototype of analysis and proactive detection of attacks related to cyberbullying was developed, which, in its precision analysis, reached values like those previously obtained. A further investigation will describe the implemented tools and functionalities.

Table 6.6: Prototype results

Theoretical Component				Practical component		
Research	Dataset	Compromise solution Topics obtained	Applied topic comparison models	AI techniques for learning	Dataset sectioning Learning - Tests	LinearSVC precision
Modeling Proposal and Prototype	Previous Grooming and Bullying Database	6 (Grooming) - 4 (Bullying)	HDP LSI LDA NMF	LinearSVC	70 % - 30 %	94 %

6.6 ANSWERING THE RESEARCH QUESTIONS

6.6.0.1 Is the proposed modeling process applicable to other cyber-attacks that evidence social engineering tactics?

The basic process started with the Grooming research. In this attack, criteria and modeling parameters were defined that were applied to a new Bullying case study. This attack was analyzed with two types of databases, unlike Bullying, and implemented in Python. The results of this study strengthened the modeling process of these attacks. For this reason, it is considered that the proposed process can model new attacks that are related to psychological manipulation. It should be noted that the process does not standardize models and tools.

6.6.0.2 Which topic model presents better results in the evaluation of short texts?

As evidenced in Section 4, the LDA model and its HDP variant currently offer better short text processing results. These texts come from instant messengers and comment blogs. It should be noted that the databases used were purified under the same cleaning and normalization process.

6.6.0.3 Can the models determined by the proposed process be implemented in functional systems today?

In Section 5, a system based on the proposed models was developed. This modular prototype allows the modification of the databases, learning models, and classification. The prototype represents the theoretical models proposed, and its operation indicates that it applies to new cyber-attacks. For this reason, currently, there are computational resources for the development of these AI-based tools.

6.7 CONCLUSIONS AND FUTURE WORK

The detection and mitigation mechanisms that counteract cyber-attacks in real-time, where the objective is human destabilization, are still under development. As evidenced in the literature review conducted, efforts have been made to eliminate the effects of these phenomena; however, they persist and manifest themselves incrementally as they improve their attack and evasion techniques. Studying these attacks directed at the human psyche from the point of view of information security, such as social engineering, makes it possible to link proposed techniques, methodologies, and architectures to future cybersecurity and conventional security projects. With this connection, it would be possible to standardize knowledge and processes on cybersecurity, thus avoiding incomplete and scattered proposals. Furthermore, the results obtained in this research consolidate the modeling process carried out on Grooming and Bullying and enable the possibility of applying it to future social engineering attacks not yet defined.

We have combined the concepts of cybersecurity, social engineering, and traditional security to understand that these cyber-attacks are part of the same line of study. With theme modeling, different stages or topics that model the attacks were determined; however, these topics require human intervention to define a lexical concept or communicational intention. This will support investigations related to identifying patterns of malicious behavior online.

In the experimentation phase, a typical attackers'pattern was determined in the processing of information related to the experiences of the victims. The statistical algorithm LDA and its predecessor HDP presented the best results in analyzing and distributing the information, delivering four groupings of words for Bullying and six for Grooming. By themselves, these

classifications do not describe linguistic aspects; therefore, linguistic software was used to define the communicational intentions of each stage. With this knowledge, the stages of the models assigned to information security were correlated, and the definitive model of our case study was defined.

REFERENCES

[1] Rossouw Von Solms and Johan Van Niekerk. From information security to cyber security. *computers & security*, 38:97–102, 2013.

[2] Patricio Zambrano, Jenny Torres, Luis Tello-Oquendo, Rubén Jácome, Marco E Benalcazar, Roberto Andrade, and Walter Fuertes. Technical mapping of the grooming anatomy using machine learning paradigms: An information security approach. *IEEE Access*, 7:142129–142146, 2019.

[3] Patricio Zambrano, Jenny Torres, Ángel Yánez, Alexandra Macas, and Luis Tello-Oquendo. Understanding cyberbullying as an information security attack—life cycle modeling. *Annals of Telecommunications*, pages 1–19, 2020.

[4] Joseph M Hatfield. Social engineering in cybersecurity: The evolution of a concept. *Computers & Security*, 73:102–113, 2018.

[5] Patricio Zambrano, Jenny Torres, and Pamela Flores. How does grooming fit into social engineering? In *Advances in Computer Communication and Computational Sciences*, pages 629–639. Springer, 2019.

[6] Neil Morelli, Denise Potosky, Winfred Arthur Jr, and Nancy Tippins. A call for conceptual models of technology in io psychology: An example from technology-based talent assessment. *Industrial and Organizational Psychology*, 10(4):634, 2017.

[7] Ricardo F Muñoz. Harnessing psychology and technology to contribute to making health care a universal human right. *Cognitive and Behavioral Practice*, 2019.

[8] Vinh-Loc Dao, Cécile Bothorel, and Philippe Lenca. Community detection methods can discover better structural clusters than ground-truth communities. In *2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 395–400. IEEE, 2017.

[9] Zubair Amin. Q methodology: A journey into the subjectivity of human mind. *Singapore medical journal*, 41(8):410–414, 2000.

[10] A Fuat Fırat and Nikhilesh Dholakia. From consumer to construer: Travels in human subjectivity. *Journal of Consumer Culture*, 17(3):504–522, 2017.

[11] Kevin D Mitnick, William L Simon, and S Wozniak. The art of deception: Controlling the human element of security. 2002. *Paperback ISBN 0-471-23712-4*, 2006.

[12] Francois Mouton, Louise Leenen, and H. S. Venter. Social engineering attack examples, templates and scenarios. *Computers and Security*, 59:186–209, 2016.

[13] Hamed Jelodar, Yongli Wang, Chi Yuan, Xia Feng, Xiahui Jiang, Yanchao Li, and Liang Zhao. Latent dirichlet allocation (lda) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications*, 78(11):15169–15211, 2019.

[14] Rania Albalawi, Tet Hin Yeap, and Morad Benyoucef. Using topic modeling methods for short-text data: A comparative analysis. *Frontiers in Artificial Intelligence*, 3:42, 2020.

[15] Rubayyi Alghamdi and Khalid Alfalqi. A survey of topic modeling in text mining. *Int. J. Adv. Comput. Sci. Appl.(IJACSA)*, 6(1), 2015.

[16] Loulwah AlSumait, Daniel Barbará, and Carlotta Domeniconi. On-line lda: Adaptive topic models for mining text streams with applications to topic detection and tracking. In *2008 eighth IEEE international conference on data mining*, pages 3–12. IEEE, 2008.

[17] Shanzeng Qiao and Aili Han. A way to construct evolution model of scientific papers based on the seed document and olda models. In *Proceedings 2013 International Conference on Mechatronic Sciences, Electric Engineering and Computer (MEC)*, pages 900–903. IEEE, 2013.

[18] Shoaib Jameel, Wai Lam, and Lidong Bing. Supervised topic models with word order structure for document classification and retrieval learning. *Information Retrieval Journal*, 18(4):283–330, 2015.

[19] Lingyun Li, Yawei Sun, and Cong Wang. Semantic augmented topic model over short text. In *2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS)*, pages 652–656. IEEE, 2018.

[20] Ding Peng, Dai Guilan, and Zhang Yong. Contextual-lda: a context coherent latent topic model for mining large corpora. In *2016 IEEE Second International Conference on Multimedia Big Data (BigMM)*, pages 420–425. IEEE, 2016.

[21] Di Liu, Ye Zeng, Yu Luo, Hong Pang, and Xiao-Hua Wu. Window-based topic model for hdp. In *2019 16th International Computer Conference on Wavelet Active Media Technology and Information Processing*, pages 70–75. IEEE, 2019.

[22] Mehdi Allahyari and Krys Kochut. Automatic topic labeling using ontology-based topic models. In *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, pages 259–264. IEEE, 2015.

[23] Huong T Le, Lam N Pham, Duy D Nguyen, Son V Nguyen, and An N Nguyen. Semantic text alignment based on topic modeling. In *2016 IEEE RIVF International Conference on Computing & Communication Technologies, Research, Innovation, and Vision for the Future (RIVF)*, pages 67–72. IEEE, 2016.

[24] Hongyu Liu, Ruifang He, Haocheng Wang, and Bo Wang. Fusing parallel social contexts within flexible-order proximity for microblog topic detection. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pages 875–884, 2020.

[25] Tian Shi, Kyeongpil Kang, Jaegul Choo, and Chandan K Reddy. Short-text topic modeling via non-negative matrix factorization enriched with local word-context correlations. In *Proceedings of the 2018 World Wide Web Conference*, pages 1105–1114, 2018.

[26] Jie Bai, Linjing Li, and Daniel Zeng. Activating topic models from a cognitive perspective. In *2016 IEEE Conference on Intelligence and Security Informatics (ISI)*, pages 55–60. IEEE, 2016.

[27] Yuan Zuo, Jichang Zhao, and Ke Xu. Word network topic model: a simple but general solution for short and imbalanced texts. *Knowledge and Information Systems*, 48(2):379–398, 2016.

[28] Rui Zhao and Kezhi Mao. Cyberbullying detection based on semantic-enhanced marginalized denoising auto-encoder. *IEEE Transactions on Affective Computing*, 8(3):328–339, 2016.

[29] Karthik Dinakar, Roi Reichart, and Henry Lieberman. Modeling the detection of textual cyberbullying. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 5, pages 11–17, 2011.

[30] Elif Varol Altay and Bilal Alatas. Detection of cyberbullying in social networks using machine learning methods. In *2018 International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism (IBIGDELFT)*, pages 87–91. IEEE, 2018.

[31] Selma Ayşe Özel, Esra Saraç, Seyran Akdemir, and Hülya Aksu. Detection of cyberbullying on social media messages in turkish. In *2017 International Conference on Computer Science and Engineering (UBMK)*, pages 366–370. IEEE, 2017.

[32] Thabo Mahlangu, Chunling Tu, and Pius Owolawi. A review of automated detection methods for cyberbullying. In *2018 International Conference on Intelligent and Innovative Computing Applications (ICONIC)*, pages 1–5. IEEE, 2018.

[33] Vikas S Chavan and SS Shylaja. Machine learning approach for detection of cyber-aggressive comments by peers on social media network. In *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 2354–2358. IEEE, 2015.

[34] Walisa Romsaiyud. Expectation-maximization algorithm for topic modeling on big data streams. In *2016 IEEE 7th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, pages 1–7. IEEE, 2016.

[35] Eren Çürük, Çiğdem Acı, and Esra Saraç Eşsiz. Performance analysis of artificial neural network based classfiers for cyberbulling detection. In *2018 3rd International Conference on Computer Science and Engineering (UBMK)*, pages 1–5. IEEE, 2018.

[36] Rob Churchill, Lisa Singh, and Christo Kirov. A temporal topic model for noisy mediums. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 42–53. Springer, 2018.

[37] Qingqiang Wu, Xiang Deng, Caidong Zhang, and Changlong Jiang. Lda-based model for topic evolution mining on text. In *2011 6th International Conference on Computer Science & Education (ICCSE)*, pages 946–949. IEEE, 2011.

[38] Seyed Ali Bahrainian, Ida Mele, and Fabio Crestani. Modeling discrete dynamic topics. In *Proceedings of the Symposium on Applied Computing*, pages 858–865, 2017.

[39] Bin Liao, Weihua Wang, and Chunmei Jia. Clustering and recommendation of scientific documentation based on the topic model. In *Proceedings of the 2012 International Conference on Information Technology and Software Engineering*, pages 629–637. Springer, 2013.

[40] Zhipeng Xie, Liyang Jiang, Tengju Ye, and Zhenying He. Mptm: A topic model for multi-part documents. In *International Conference on Database Systems for Advanced Applications*, pages 154–168. Springer, 2015.

[41] Amine Trabelsi and Osmar R Zaïane. A joint topic viewpoint model for contention analysis. In *International Conference on Applications of Natural Language to Data Bases/Information Systems*, pages 114–125. Springer, 2014.

[42] Xueqi Cheng, Xiaohui Yan, Yanyan Lan, and Jiafeng Guo. Btm: Topic modeling over short texts. *IEEE Transactions on Knowledge and Data Engineering*, 26(12):2928–2941, 2014.

[43] Aytuğ Onan. Two-stage topic extraction model for bibliometric data analysis based on word embeddings and clustering. *IEEE Access*, 7:145614–145633, 2019.

[44] Li Sun, Jie Chen, Jiyun Li, and Yingli Peng. Joint topic-opinion model for implicit feature extracting. In *2015 10th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*, pages 208–213. IEEE, 2015.

[45] Fan Zhang, Wang Gao, Yuan Fang, and Bo Zhang. Enhancing short text topic modeling with fasttext embeddings. In *2020 International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)*, pages 255–259. IEEE, 2020.

[46] Zheng Liu, Tingting Qin, Ke-Jia Chen, and Yun Li. Collaboratively modeling and embedding of latent topics for short texts. *IEEE Access*, 8:99141–99153, 2020.

[47] Wenxin Liang, Ran Feng, Xinyue Liu, Yuangang Li, and Xianchao Zhang. Gltm: A global and local word embedding-based topic model for short texts. *IEEE access*, 6:43612–43621, 2018.

[48] Lingyun Li, Yawei Sun, Xu Han, and Cong Wang. Research on improve topic representation over short text. In *2018 IEEE Third International Conference on Data Science in Cyberspace (DSC)*, pages 848–853. IEEE, 2018.

[49] Mingyang Xu, Ruixin Yang, Stephen Ranshous, Shijie Li, and Nagiza F Samatova. Leveraging external knowledge for phrase-based topic modeling. In *2017 Conference on Technologies and Applications of Artificial Intelligence (TAAI)*, pages 29–32. IEEE, 2017.

[50] Ximing Li, Changchun Li, Jinjin Chi, and Jihong Ouyang. Short text topic modeling by exploring original documents. *Knowledge and Information Systems*, 56(2):443–462, 2018.

[51] Changsong Dai, Yongbin Wang, and Qi Wang. Topic model and similarity calculation of text on spark. In *2017 14th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, pages 15–19. IEEE, 2017.

[52] Sandhya Subramani, Vaishnavi Sridhar, and Kaushal Shetty. A novel approach of neural topic modelling for document clustering. In *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 2169–2173. IEEE, 2018.

[53] Xiaobao Wu and Chunping Li. Short text topic modeling with flexible word patterns. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–7. IEEE, 2019.

[54] BS Jadhav, DS Bhosale, and DS Jadhav. Pattern based topic model for data mining. In *2016 International Conference on Inventive Computation Technologies (ICICT)*, volume 2, pages 1–6. IEEE, 2016.

[55] Peng Yang, Wenhan Li, and Guangzhen Zhao. Language model-driven topic clustering and summarization for news articles. *IEEE Access*, 7:185506–185519, 2019.

[56] Jiamiao Wang, Ling Chen, Lu Qin, and Xindong Wu. Astm: An attentional segmentation based topic model for short texts. In *2018 IEEE International Conference on Data Mining (ICDM)*, pages 577–586. IEEE, 2018.

[57] Hosein Azarbonyad, Mostafa Dehghani, Tom Kenter, Maarten Marx, Jaap Kamps, and Maarten De Rijke. Hitr: Hierarchical topic model re-estimation for measuring topical diversity of documents. *IEEE Transactions on Knowledge and Data Engineering*, 31(11):2124–2137, 2018.

[58] Fei Wang, Rui Liu, Yuan Zuo, Hui Zhang, He Zhang, and Junjie Wu. Robust word-network topic model for short texts. In *2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 852–856. IEEE, 2016.

[59] Than Than Wai and Sint Sint Aung. Enhanced frequent itemsets based on topic modeling in information filtering. *International Journal of Software Innovation (IJSI)*, 5(4):33–43, 2017.

[60] Hyungsul Kim, Yizhou Sun, Julia Hockenmaier, and Jiawei Han. Etm: Entity topic models for mining documents associated with entities. In *2012 IEEE 12th International Conference on Data Mining*, pages 349–358. IEEE, 2012.

[61] Shuangyin Li, Yu Zhang, and Rong Pan. Bi-directional recurrent attentional topic model. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 14(6):1–30, 2020.

[62] Di Jiang, Yuanfeng Song, Yongxin Tong, Xueyang Wu, Weiwei Zhao, Qian Xu, and Qiang Yang. Federated topic modeling. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pages 1071–1080, 2019.

[63] Thuc Nguyen and Phuc Do. Citationlda++ an extension of lda for discovering topics in document network. In *Proceedings of the Ninth International Symposium on Information and Communication Technology*, pages 31–37, 2018.

[64] Noriaki Kawamae. Topic chronicle forest for topic discovery and tracking. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pages 315–323, 2018.

[65] Yang Yang, Feifei Wang, Fei Jiang, Shuyuan Jin, and Jin Xu. A topic model for hierarchical documents. In *2016 IEEE First International Conference on Data Science in Cyberspace (DSC)*, pages 118–126. IEEE, 2016.

[66] Wenxing Hong, Xiaoqing Zheng, Jianwei Qi, Weiwei Wang, and Yang Weng. Project rank: An internet topic evaluation model based on latent dirichlet allocation. In *2018 13th International Conference on Computer Science & Education (ICCSE)*, pages 1–4. IEEE, 2018.

[67] Qiuxing Chen, Lixiu Yao, and Jie Yang. Short text classification based on lda topic model. In *2016 International Conference on Audio, Language and Image Processing (ICALIP)*, pages 749–753. IEEE, 2016.

[68] Tinh Thanh Dao, Tinh Dao Thanh, Thanh Nguyen Hai, and Vinh Ho Ngoc. Building vietnamese topic modeling based on core terms and applying in text classification. In *2015 Fifth International Conference on Communication Systems and Network Technologies*, pages 1284–1288. IEEE, 2015.

[69] Seonggyu Lee, Jinho Kim, and Sung-Hyon Myaeng. An extension of topic models for text classification: A term weighting approach. In *2015 International Conference on Big Data and Smart Computing (BIGCOMP)*, pages 217–224. IEEE, 2015.

[70] Hongchen Guo, Qiliang Liang, and Zhiqiang Li. An improved ad-lda topic model based on weighted gibbs sampling. In *2016 IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*, pages 1978–1982. IEEE, 2016.

[71] Dexin Zhao, Jinqun He, and Jin Liu. An improved lda algorithm for text classification. In *2014 International Conference on Information Science, Electronics and Electrical Engineering*, volume 1, pages 217–221. IEEE, 2014.

[72] Scott Deerwester, Susan T Dumais, George W Furnas, Thomas K Landauer, and Richard Harshman. Indexing by latent semantic analysis. *Journal of the American society for information science*, 41(6):391–407, 1990.

[73] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022, 2003.

[74] Daniel D Lee and H Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999.

[75] Yee Whye Teh, Michael I Jordan, Matthew J Beal, and David M Blei. Hierarchical dirichlet processes. *Journal of the american statistical association*, 101(476):1566–1581, 2006.

[76] José Rúas Araújo, Mercedes Fernández Cabana, and Iván Puentes Rivera. Aplicación de la herramienta liwc al análisis del discurso político: Los mítines de los candidatos en las elecciones al parlamento de galicia de 2012. In *Investigar la Comunicación Hoy. Revisión de políticas científicas y aportaciones metodológicas: Simposio Internacional sobre Política Científica en Comunicación*, pages 47–64. Facultad de Ciencias Sociales, Juridicas y de la Comunicación, 2013.

[77] Ethan Fast, Binbin Chen, and Michael S Bernstein. Empath: Understanding topic signals in large-scale text. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 4647–4657, 2016.

7 DISCUSSION

This chapter presents to what extent the proposed process supported the solution of the research problem. This was evidenced by contrasting the objectives established and the results obtained. In the testing phase, a typical pattern of cyberbullies in the processing of information related to the experiences of the victims was determined. The statistical and computational algorithm LDA (Latent Dirichlet Allocation) classified all the information into four themes or stations that were part of the cyberbullying model.

By themselves, these classifications did not describe linguistic aspects; therefore, linguistic software was used to define the communicational intentions of each stage. With this knowledge, the stages of the models assigned to the field of information security were correlated, and the definitive model of our case study was defined.

This evaluation was obtained through empirical evidence and feedback from experts and specialists in the field, through peer review journals. This approach facilitated a better understanding of the problem.

Understanding that the problem is developed because of the lack of a formal processes that allow the study of online harassment, a series of interlinked contributions were made to provide a formal and procedural solution to this. The linkage of the contributions are described below:

- The results of conference 1, Cyber Security in Networking Conference (CSNet), made it possible to understand and clarify the problems associated with online pedophilia, grooming as an attack technique, and the contributions found in the scientific field.
- The knowledge acquired in the research process related to conference 2, Advances in Intelligent Systems and Computing, established the link of grooming within the field of social engineering.
- With these results and a further analysis of the literature, it was possible to create a

psychological/technical profile of cyber attackers, and the life cycle of the attack based on 6 stages, which are aligned with pre-established concepts of information security (APTs). These results were evaluated and approved in the journal IEEE Open Access (JCR-Q1).

- In the fourth contribution, an analysis of the input data was carried out, with download and debugging techniques proposed in the field of text mining. With these data and the process applied in the evaluation of grooming, a transferability to another phenomenon of harassment, called cyberbullying, was made. The results obtained were published in the Annals of Telecommunications journal (JCR-Q3).

- Finally, in the fifth research in the Journal of Information Security and Applications (JCR-Q2), which is in the second round of evaluation, the proposed process was formalized, and a parental control prototype was described as an innovation complement to the final research.

Below, we present a discussion between the research questions and the contributions presented in this thesis in order to analyze how they contribute in solving the problem proposed.

RQ1. *What is the conceptual and procedural relationship between grooming and social engineering?*

Different researches related to the study of conventional computer security attacks, outlines a life cycle, defining the events that occur in each stage. Since this is a chronological process that takes place during a computer attack from its beginning, until the achievement of its objective. Kevin Mitnick proposed 4 phases of the life cycle of attacker behavior when applying social engineering: 1) information gathering, 2) development of relationship, 3) exploitation of relationship, and 4) execution to achieve objective.

There are contributions related to studies of more advanced attacks known as Advanced Persistent Threat (APT), that come from highly prepared, organized, sophisticated, determined, and obstinate cybercriminals who direct their attacks to specific people or organizations, government institutions, among others, with the purpose of obtaining competitive advantages and strategic benefits that in some situations cause irreparable damage.

The success of these attacks is based on repetition, maintaining discretion and immediate non-invasiveness, but with high resistance capabilities, in the long term to meet their objec-

tives. In the scientific field, social engineering is considered a type of APT, since attackers in their behavioral patterns demonstrate similar aspects such as persistence, discretion, attack execution time, among others. With the theoretical results obtained (Contributions 1 and 2, see Section 1 in Figure 7.1) when analyzing the structure of grooming, the behavior of the attackers and their scope, it was evidenced its direct relationship with the processes of social engineering.

In the first instance, these were linked to the 4 stages proposed by Mitnick. Thus, these attacks work as follow: *1) collection of information from the perpetrators*, *2) development of the relationship with their victims* applying tactics such as persuasion, physical encounters, and the application of new technologies with the intention of not leaving a digital trail, among others. Regarding the third phase, *3) exploitation of the relationship*, the aggressors have greater control over their victims, thus achieving isolation, sexual desensitization, personal information about them, etc., y *4) when achieving their objectives*, the attacker can obtain anything from the victim, like pornographic material or even physical encounters.

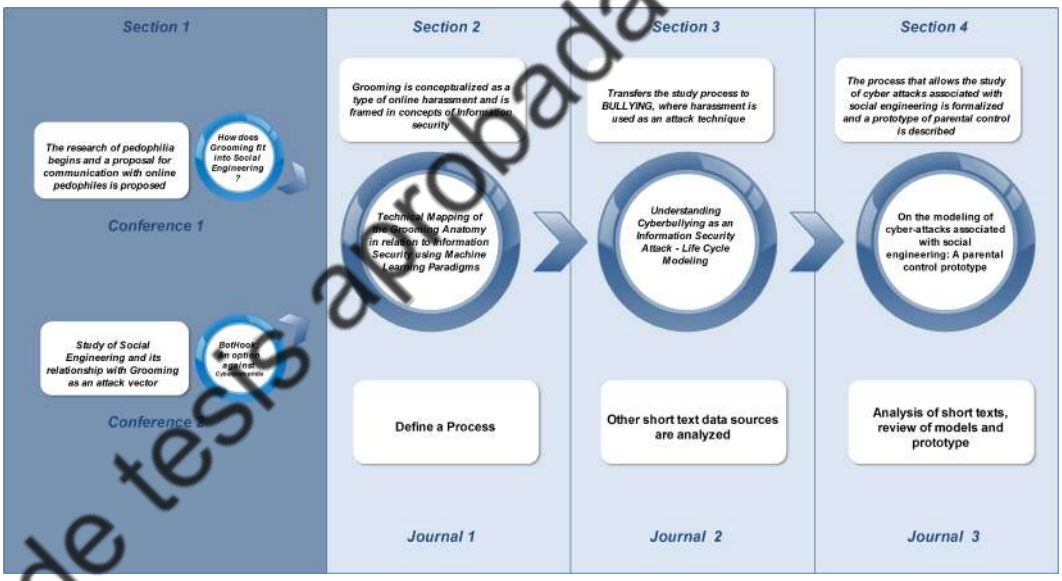


Figure 7.1: Contributions - Section 1.

In relation to grooming studies, the literature shows that it is carried out in several stages (life cycle), however the results of the research differ from each other, since they have been supported by psychological criteria with a high degree of subjectivity. In this context, this research presented its results, see Section 2 in Figure 7.2, based on the application of an Latent Dirichlet Allocation (LDA) statistical model that allowed determining 6 stages with a lower degree of subjectivity.

With the criteria indicated and supported in relation to the study of APTs, we established

the direct correlation between grooming as an attack vector of social engineering in conceptual aspects. In procedural aspects, the research confirms that grooming follows a life cycle, which develops in 6 stages: 1) information gathering, 2) gaining access, 3) lateral movement, 4) privilege escalation, 5) execution to achieve the objective and 6) reporting.

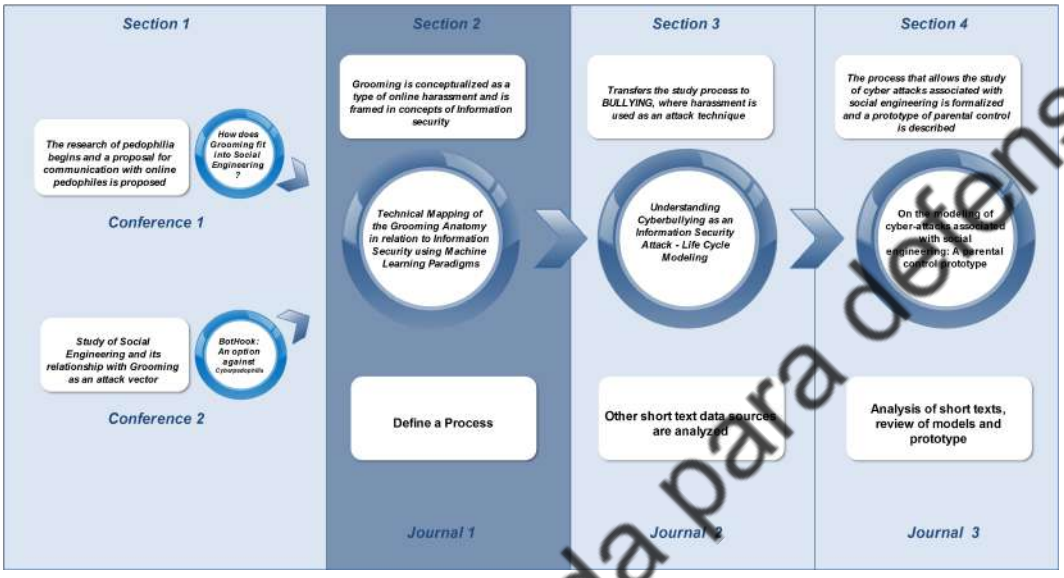


Figure 7.2: Contributions - Section 2.

RQ2. *What are the theoretical foundations provided by the study of grooming in the phenomenon associated with cyberbullying?*

This study of grooming allowed to establish a life cycle with 6 phases, where we established the mechanisms applied in each one of them towards the victims. This research was based on the analysis of short texts taken from the website Perverted Justice (PJ). However, in an extension of the research, cyberbullying was analyzed as a second phenomenon related to harassment, see Section 3 in Figure 7.3. From the analysis of the literature, it was possible to see that there are greater technological contributions of this phenomenon in contrast to grooming, however, it is not supported by concepts of social engineering, standard procedures to determine phases or data acquisition that allow a better context.

Under these considerations, when carrying out the study, there was not a single database since it was based on tweets of short texts and experiences of the victims. Thus, the analysis of richer texts (experiences) was necessary, since the tweets obtained (very scattered data) did not allow a deeper analysis of the process by which victims are harassed. The structured information of the experiences was better adapted to the modeling of topics, establishing 4

very well defined themes: *1) development of relationship, 2) preparing/distracting attack, 3) exploit the relationship and 4) debrief.*

The way cyberbullying is applied to victims differs from the grooming life cycle in 2 phases. This is understandable since the experiences describe a more superficial process of the attack, which is not the case in PJ chats. However, the two phenomena clearly describe the phases by which the victims are harassed with techniques of persuasion, persistence and preparation. These phases consolidate social phenomena such as grooming and cyberbullying within the attack vectors of social engineering.

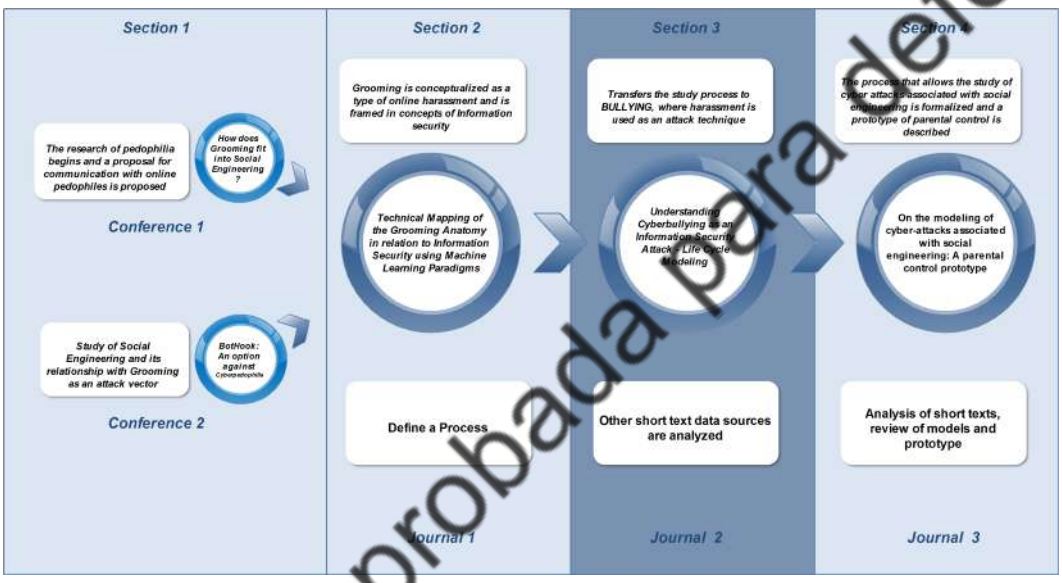


Figure 7.3: Contributions - Section 3.

RQ3. *How the different types of online harassment are related in their respective contexts or phenomena?*

In the development of the study of grooming, a procedural mechanism was established for its analysis, which requires 4 aspects to be transferred to other phenomena, see Section 4 in Figure 7.4. Within these aspects, we have:

1. **To describe the type of data to be analyzed.** Here it is necessary to verify the existence of previous studies, or proposals that present data download procedures or specific modeling. In this aspect, researchers must collect texts, chats or any information related to instant messaging. It is worth mentioning that these data must be related to the use of psychological manipulation techniques or social pressure by the attackers for the research to be viable.

2. **To select a data mining methodology.** It is recommended to apply CRISP-DM as a methodology, since in this way the data can be cleaned for subsequent treatment.

3. **To define a topic modeling mechanism.** The use of topic modeling mechanisms is for the analysis of refined data. It should be noted that the precision of these models is not within the scope of this research.

4. **To consider a linguistic context.** It is necessary because the results of the models do not have an explicit meaning. With the use of software tools or manual dictionary-based categorization, lexical categories and communication intentions are defined. This information is contrasted with models associated with the APTs and a new life cycle is obtained.

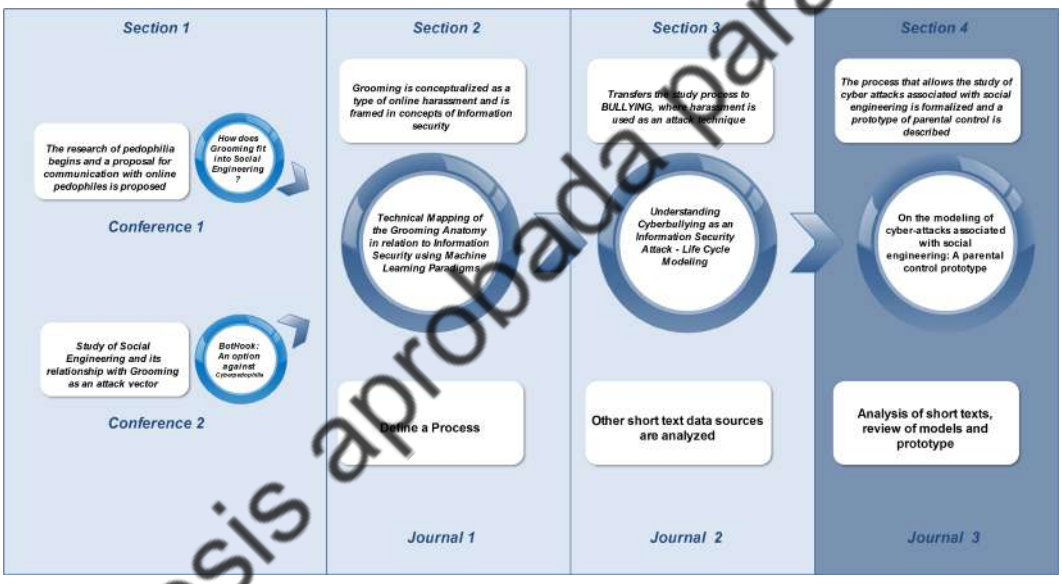


Figure 7.4: Contributions - Section 4.

In summary, the types of online harassment are related to the use of communication techniques to harass people, the media or text technologies, potential victims with psychological problems such as low self-esteem, depression, and other emotional aspects that can be used by stalkers. These aspects have been observed in the research that support this thesis.

8 CONCLUSIONS

The research consider two aspects of study. The first aspect is a theoretical aspect that allowed to answer the research questions raised in the study of grooming and cyberbullying. The second aspect contemplates the application part of the project. Table 8.1 summarizes the results of the study aspects.

Table 8.1: Summary of previous research

Theoretical Component						Research Application		
Research	Dataset		Average of topics proposed in the literature	Compromise solution Topics obtained	Applied models to get topics	AI techniques for learning	Dataset sectioning to learning & Tests	Model accuracy result
grooming	100 conversations 128171 chat lines		6	6	LDA	LinearSVC CNN LSTM	90 % - 10 %	97,61 % 96,11 % 95,91 %
cyberbullying	250,000 attacker related tweets	3035 victim experiences	-	4	PLSA LDA	LinearSVC LSTM	70 % - 30 %	94,74 % 94,41 %

The research includes different fields related to information security, as is the case of social engineering. The processing and evaluation of short instantaneous texts, through the proposed approach, not only applies to the case study, but can be transferred to other fields related to the behavioral patterns of people before the application of computer security in their lives. One of the main challenges in new case studies is obtaining relevant data related to the field of research; therefore, it is essential to promulgate and disseminate this type of study in the scientific community to gain more interest in academia and industry.

With the development of the research, the results obtained from it and the approach of a procedural process, grooming was positioned as an attack associated with social engineering and information security. Through the definition of the phases associated with the life cycle of grooming and cyberbullying; this will allow to support research related to the identification of patterns of malicious online behavior.

As can be seen, the study of unconventional attacks related to online harassment has several aspects that are of social and scientific interest. With the support of systems and technology, psychological phenomena can be studied that would not be feasible with physical

participants.

With the purpose of establishing a general framework in studies related to the human behavior of cybercriminals, empirical studies have tried to define phases within the life cycle of attacks. Kevin Mitnik in [11] from the perspective of social engineering established 4 phases. In the study of O'Connell, regarding grooming, he established 6 phases which were accepted and approved in other research [19, 62]. Other studies proposals related to APT are more diverse and range from 3 to 8 phases [72].

Our results show that the life cycle related to grooming has 6 phases associated with concepts of social engineering. This number of topics confirms what was proposed by O'Connell. However, our process applied CRISP-DM as a data methodology for data analysis and topic modeling. Making this process a transferable resource to other phenomena, such as the case of cyberbullying, which has not been evaluated from these perspectives.

8.1 THEORETICAL ASPECTS

Detection and mitigation mechanisms that counteract unconventional attacks, where the means of communication is human interaction, are still under development. As evidenced in the literature review, efforts have been made to eliminate the effects of this phenomenon; however, it persists and manifests itself more severely in its victims. Studying this phenomenon from the point of view of information security, such as social engineering, allows linking proposed techniques, methodologies, and architectures to online harassment and viceversa. With this link, it would be possible to develop more specialized knowledge on the subject, thus avoiding incomplete and scattered proposals. The results obtained in this research justify the validity of the process of obtaining unconventional attack models, as evidenced in a previous research related to grooming and cyberbullying.

Grooming as an access technique has shown in our research that it focuses on an extremely sensitive group in society, children, and adolescents. It is the most dangerous technique in information security since it not only attacks material assets but also the emotional and mental stability of people, which could have consequences for life. Computer science, in its continuous contribution to society, is establishing new aspects of social engineering, for a better understanding and research of this phenomenon. We established the interrelationship between social engineering and grooming. Our mail goal was to position and promote the

study of grooming, since we consider that it must be part of a taxonomy, thus generating a significant contribution to the development of previously established knowledge.

8.2 PRACTICAL ASPECTS

Income data

In the development of the research, it became evident that the use of Perverted-Justice databases in studies related to online Pedophilia was common. However, the research not only wanted to stay in this context and broadened the field of study to other short text data and even analyzed more enriched text such as experiences of victims (experience blogs) for the case of cyberbullying, through the transferability of the proposed study process. All these data allowed the study to be conducted from a technological perspective with the use of text mining techniques and machine learning techniques.

Topic modeling

The data was preprocessed applying the CRISP-DM methodology and analyzed with topic modeling. In the case of grooming, these results were contrasted with the theoretical results. These results, the number of topics, were obtained from the perplexity and computational cost analysis, defining a compromise solution that determined 6 stages for grooming and 4 stages for cyberbullying in the respective life cycles of each attack. Within the application of the modeling, the communicational intentions were defined with linguistic concepts to the established topics.

Machine learning

In a comparative analysis of the precision of the results, the models applied in the literature have a lower precision than those obtained in our research, see Table 8.2. It is worth mentioning that elements such as: selection of n-grams, sectioning of texts for learning and text refinement processes are aspects that contributed to improve the results in the laboratory.

By evaluating the prototype developed with the number of topics, the classification of the text according to the phases, linguistic characteristics, a percentage of data for learning and the application of the linear model of automatic learning, an accuracy of 94 % was evidenced when classifying unknown data by the system, see Table 8.3.

Table 8.2: Related research precision measurements

Ítem	Technique used	Technique accuracy	Stages	Reference
Research 1	SVM (Support Vector Machines)	97 %	4	[4]
Research 2	Naive Bayes classification	94 %	4	[27]
Research 3	SVM and k-NN (K Nearest Neighbor)	97.08 %	6	[62]
Research 4	Logistic models	95 %	6	[19]
Research 5	SVM and k-NN	94.3 %	-	[2]
Research 6	Naive Bayes	96 %	-	[18]
Research results	Linear SVC (Support Vector Classifier)	97.61 %	6	

Table 8.3: Prototype results

Theoretical Component				Practical component		
Research	Dataset	Compromise solution Topics obtained	Applied topic comparison models	AI techniques for learning	Dataset sectioning Learning - Tests	LinearSVC precision
Modeling Proposal and Prototype	Previous Grooming and Cyberbullying Database	6 (Grooming) - 4 (Cyberbullying)	HDP LSI LDA NMF	LinearSVC	70 % - 30 %	94 %

8.3 METHODOLOGICAL ASPECTS

The qualitative method selected to develop conceptual and procedural aspects, allowed obtaining results that, in contrast to the findings in the literature, enabled a deeper understanding of a social phenomenon that continues growing. In addition to what was described above, this method allowed identifying the necessary steps to solve the determined research problem, so it can be concluded that it is applicable to future studies in the field of cybersecurity. Besides the scientific method that conducted the research, the methodology of CRISP-DM was used in order to establish the formal and methodological process that studies grooming in relation to the criteria and concepts of computer security.

CRISP-DM is a comprehensive data mining methodology and process model that provides a comprehensive blueprint for conducting a data mining project. CRISP-DM breaks down the life cycle of a data mining project into 6 phases: 1) understanding the problem, 2) understanding the data, 3) preparing the data, 4) modeling, 5) evaluating, and 6) implementing. This model points out the most frequent dependencies between the phases, as well as the cyclical nature of data mining, and illustrates that the lessons learned during the data extraction process and the implemented solution can trigger new, usually more specific, research questions [10, 11].

8.4 FUTURE WORK

As future work, from a theoretical point of view, it is proposed to analyze improvements to the LDA model including a framework of communicative intention to obtain better precision in the definition of topics when analyzing short texts. Another aspect to be considered is the study of other similar phenomena, but with more enriched texts such as blogs or experiences of the victims and contrast them with the results obtained. In practical aspects, the implementation in real environments of intelligent and proactive applications is essential. In this context, it is proposed to develop applications specific of the language to be studied.

In addition, establishing stages of analysis of short audios sent by instant messaging, as well as the study of emoticons and symbols that establish malicious communicative intentions will allow for refining the processes established in the thesis.

The proposed prototype has certain limitations such as the use of an exclusive API for the Google Chrome browser; also the use of a specific machine learning technique; it was developed exclusively for the Windows operating system; and finally that it does not support mobile devices. From the technical point of view, the aforementioned limitations can be taken as case studies or gaps for its adaptation and scalability to new technological trends.

From a societal point of view, this thesis opens a space for learning, aimed at minors, about the existing psychological attacks on the Internet. This learning can take place in controlled spaces with professional guidance. In these spaces, it is possible to evaluate the response of minors to conversations that seek to obtain confidential information from them. In this way, their awareness of the use of their information in digital environments is developed.