Joakim Granli Antonsen

Cyber Grooming Detection

Human or Machine? Or Hybrid?

Master's thesis in Information Security
Supervisor: Patrick Bours
December 2021

NTNU
Norwegian University of
Science and Technology

Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Dept. of Information Security and Communication Technology

Joakim Granli Antonsen

Cyber Grooming Detection

Human or Machine? Or Hybrid?

Master's thesis in Information Security
Supervisor: Patrick Bours
December 2021

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Dept. of Information Security and Communication Technology

NTNU
Norwegian University of
Science and Technology

Cyber Grooming Detection: Human or Machine? Or Hybrid?

Joakim Granli Antonsen

December 14, 2021

Abstract

The technological evolution is providing new opportunities at fast pace. Such opportunities also come with new challenges, one of which is cyber grooming. Predators are taking advantage of the new opportunities being present everywhere, establishing relations to children. The relations are further misused by the predators to perform sexual abuse or other malicious actions. This thesis aims to discover extended knowledge of features found from human analysis of predatory and non-predatory chat conversations. Features of which potentially can be used to improve cyber grooming detection systems. There exist cyber grooming detection systems utilizing machine learning algorithms, but machine learning algorithms can only discover so much on their own. Human evaluations of conversations were collected. The collected evaluations and corresponding conversations were further analyzed to discover trends, patterns and features of defining nature. No feature stood out as absolute in every predatory conversation, meaning one feature alone cannot with absolute certainty tell if a conversation is predatory or non-predatory. Combinations of two or more features were, however, found to almost always be present in predatory conversations. Several features showed to be of a defining nature. Some features are age defining and others are defining potentially intended actions. Non-predatory conversations constituted most of the evaluations, mostly being defined as normal conversations and some being sexual conversations. In order to potentially make use of the features, implementations of various machine learning methods can be included in existing cyber grooming detection systems, as for example AiBA (Author input Behavioral Analysis). Features can add to systems in different ways in order to improve detection and help detect ongoing cyber grooming at an earlier point of time.

Sammendrag

Den teknologiske utviklingen byr på mange nye muligheter. Med slike muligheter kommer også nye utfordringer. En av disse er cybergrooming. Predatorer utnytter disse nye mulighetene som lar dem være til stede over alt i det digitale rom, for å etablere relasjoner til barn. Disse relasjonene blir videre misbrukt av predatorene for å utføre seksuelle overgrep eller andre straffbare handlinger. Denne oppgaven tar sikt på å oppdage utvidet kunnskap om egenskaper fra menneskelig analyse, som mennesker finner definerende ved predator-samtaler og ikke-predator-samtaler. Egenskaper som potensielt kan benyttes til å forbedre cybergrooming-deteksjonssystemer. Det eksisterer cybergrooming-deteksjonssystemer som benytter seg av maskinlæringsalgoritmer, men det er begrenset hvor mye maskinlæringsalgoritmer klarer å lære seg på egen hånd. Det ble samlet inn menneskelige evalueringer av samtaler. Disse evalueringene ble sammen med de korresponderende samtalene videre analysert for å oppdage trender, mønstre og egenskaper av definerende art. Ingen egenskaper sto seg frem som absolutte i alle predator-samtaler, noe som betyr at en egenskap alene ikke med absolutt sikkerhet kan si om en samtale potensielt er en predator-samtale eller ikke. Kombinasjoner av to eller flere egenskaper ble funnet til å stort sett alltid være til stede i predator-samtaler. Flere egenskaper viste seg å være av definerende art. Noen egenskaper er aldersdefinerende, og andre definerer potensielt tiltenkte handlinger. Ikke-predator-samtaler utgjorde mesteparten av evalueringene, og de fleste av disse var definert som normale samtaler. Noen samtaler var også definert som seksuelle. For å potensielt kunne benytte disse egenskapene kan implementasjoner av ulike maskinlæringsmetoder inkluderes i eksisterende cybergrooming-deteksjonssystemer, som for eksempel AiBA (Author input Behavioral Analysis). Egenskaper kan legges til i systemer på ulike måter for å forbedre deteksjon og hjelpe med deteksjon av pågående cybergrooming på et så tidlig tidspunkt som mulig.

Preface

This thesis is the final project of my Master of science in Information Security at the Department of Information Security and Communication Technology, Norwegian University of Science and Technology (NTNU). The work has been supervised by responsible professor Patrick Bours.

Joakim Granli Antonsen
Oppdal, December 14, 2021

Acknowledgements

I would like to thank my supervisor Professor Patrick Bours at the Department of Information Security and Communication Technology, Norwegian University of Science and Technology (NTNU). Patrick did an outstanding job as my supervisor, asking the right questions when needed and motivating me throughout the process. Whenever I was in need of guidance or had questions, he was always available and provided the help needed.

I would also like to thank all the participants in the data collection experiment for contributing to my research.

Last but not least, I would like to thank my parents, Unn and Morten, and my brother, Erlend. Thank you for always supporting and encouraging me towards new goals, and for giving me the opportunity to write this thesis from one of my favorite places in the world, Oppdal.

Joakim Granli Antonsen
Oppdal, December 14, 2021

Contents

Abstract . . . iii

Sammendrag . . . v

Preface . . . vii

Acknowledgements . . . ix

Contents . . . xi

Figures . . . xv

Tables . . . xvii

1 Introduction . . . 1

1.1 Topic Covered by the Thesis . . . 1

1.2 Keywords . . . 2

1.3 Problem Description . . . 2

1.4 Justification, Motivation and Benefits . . . 3

1.5 Research Questions . . . 3

1.6 Planned Contributions . . . 4

2 Background . . . 5

2.1 Grooming . . . 5

2.1.1 Cyber Grooming . . . 6

2.2 Machine Learning . . . 8

2.2.1 Classification . . . 10

2.2.2 Regression . . . 10

2.2.3 Data Preprocessing . . . 11

2.3 Natural Language Processing (NLP) . . . 11

2.3.1 Bag of Words . . . 12

2.3.2 TF-IDF . . . 12

2.3.3 Word Embedding . . . 13

2.4 Statistics . . . 14

2.4.1 Precision, Recall and F-score . . . 14

2.5 Cyber Grooming Detection . . . 15

3 Data . . . 21

3.1 PAN-2012 . . . 21

3.1.1 Hybrid . . . 22

3.1.2 Dataset Structure and Features . . . 23

3.2 Data Collection from Experiment . . . 24

3.2.1 Participants . . . 24

3.2.2 The Experiment 26
3.2.3 Data Result from Experiment 27
3.3 Data Preparation 28
4 Analysis and Results 33
4.1 Analysis Method 33
4.2 Evaluations and Conversations 34
4.3 Hybrid 36
4.3.1 True Positives Hybrid 36
4.3.2 False Positives Hybrid 37
4.3.3 True Negatives Hybrid 40
4.3.4 False Negatives Hybrid 43
4.4 PAN 45
4.4.1 True Positives PAN 45
4.4.2 False Positives PAN 50
4.4.3 True Negatives PAN 53
4.4.4 False Negatives PAN 56
4.5 Most Prominent Features 58
5 Discussion 61
5.1 Age 61
5.1.1 Age Given in Chat 62
5.1.2 Impersonation 63
5.1.3 Age Not Stated 64
5.1.4 Daily Life 64
5.1.5 Living Situation 65
5.1.6 Allowed, Not Allowed or Have To 66
5.2 Meeting 67
5.3 Attention and Driving Force 67
5.3.1 Nice Words and Nicknames 67
5.4 Sexual Conversations 68
5.5 Secrecy 70
5.6 Normal Conversations 70
5.7 Consistency in Features 71
5.8 How to Make a System Hybrid 72
5.8.1 Detection Based on Questions Asked 73
5.8.2 Age Detection 73
5.8.3 Sexual Conversation Detection 74
5.8.4 Normal Conversation Detection 74
5.8.5 Feature Summary 74
5.9 Disadvantages and Experiences 75
5.9.1 Dataset 75
5.9.2 Data Collection Experiment 75
6 Conclusion and Future Work 77
6.1 Conclusion 77
6.2 Future Work 78

Bibliography **79**

Figures

2.1 Message score and risk development throughout a conversation using AiBA 18
2.2 Total risk development graph in AiBA 19

3.1 Properties of the original PAN-2012 dataset collected by [31] 22
3.2 Structure of XML file. 25
3.3 The graphical user interface of the experiment (GUI) 27
3.4 Experiment GUI: A conversation is marked as non-predatory and sexual. A few words is to be written before the submission. 28
3.5 Experiment GUI: The remaining messages of the conversation is displayed to the user after submission. 29
3.6 Excerpt from the CSV file generated by the python script. 31

4.1 Example of a predator open about age, turning the conversation sexual and initiates meeting. 38
4.2 Example of a non-predatory normal conversation between two young people getting to know each other by asking questions and sharing information. 42
4.3 Example of a sexual, predatory conversation evaluated false negative where the two conversation parties seems to know each other from before. 44
4.4 Example of a predator knowing what he is doing is wrong. 47
4.5 Example of a predator turning the conversation sexual. 49
4.6 Example of a victim introducing information about living situation. 51
4.7 Normal conversation between two chatters exchanging basic information. 55
4.8 Excerpt: Predator leading conversation, turning it sexual and asking more questions. The victim reveals information making it possible to assume age. 59

Tables

2.1 Confusion matrix 14

4.1 Number of evaluations on the different datasets. 35
4.2 Characteristics of the evaluated conversations based on the known ground truth for the different datasets. 35
4.3 Confusion matrix representing characteristics of the evaluations on the Hybrid dataset 36
4.4 Confusion matrix representing the evaluations on the PAN dataset . 45

Chapter 1

Introduction

1.1 Topic Covered by the Thesis

The world we live in is constantly developing in many ways, but the most significant and impactful is probably the technological development. Technology allows for new ways of doing things, as well as new opportunities and challenges. One such challenge introduced due to new opportunities, is cyber grooming. Cyber grooming could be stated to be the extension of grooming, utilizing new technology, where an adult wants to establish relations to children or minors online. The goal of this action is for the adult person, which we also refer to as a predator, to arrange meetings for performing sexual abuse, get nude pictures, perform sexual actions on webcam, or other malicious actions [1]. This is a very serious concern exposing the children of the world to potential sexual abuse or other inhuman, malicious actions that could create both physical and mental wounds for life [2].

Cyber grooming is extremely important to handle. As the use of internet and general awareness about cyber grooming has increased, so has the focus on detection of it as an area of research. Most research have been focused on detection based on complete chat conversations, but in the later years the focus has also been targeting live detection as early as possible in order to be able to prevent unwanted situations [3, 4].

By utilizing machine learning models, research has provided ways to perform live detection of cyber grooming with good accuracy. However, these models are trained on datasets from chat logs where the conversations of assumed predatory art are labeled. This method makes it possible for the machine learning algorithm to find patterns on its own in order to define a model to recognize potential predatory conversations [3]. Obviously, this makes the detection somewhat limited to the machine learning models ability to decide on what is considered as innocent or dangerous in a conversation. This is where the human mind is outstanding and brilliant, as human beings often easily can decide if messages in a conversation seem to be innocent or dangerous based on experience.

In this project, we will look further into what defining features can be found from human analysis, and if such potentially can be used to improve existing cyber grooming detection systems to perform detection at an even earlier point of time.

1.2 Keywords

Cyber grooming, cyber grooming detection, human analysis, machine learning, natural language processing.

1.3 Problem Description

Successful execution of cyber grooming can potentially lead up to the cause of extreme harm to individuals exposed to it. The purpose of cyber grooming is for an adult person to establish trust and build a relationship with a child or minor. This is further misused by the adult person in order to perform sexual abuse or other malicious actions.

In order to avoid sexual abuse and other malicious actions as a result of cyber grooming, it is essential to detect it, and it should be detected as early in the conversation as possible. By such detection it is possible to issue a warning to the potential victim of the chat. A human moderator at the chat provider can also be warned about a potential predatory conversation going on. After manual review by the moderator, based on the severity of the conversation, it can for example be closed, the potential predatory user can be blocked out, and law enforcement agencies can be warned.

There exists machine learning models for continuous live detection of cyber grooming in online, one-on-one conversations. The live detection works well and can perform detection at a relatively early stage. However, it should preferably perform better, detecting dangerous conversations at an even earlier stage, because time is essential. This is not very easy, because conversations can for a long time be just ordinary talk without any signs of obvious grooming. Even though it is just normal talk, the creation of a relation is in progress as messages exchanges back and forth.

When reading conversations, human beings have different prerequisites than machine learning detection models for detecting if it is dangerous and potentially predatory. In some cases humans can detect a predatory conversation after just a few messages where a machine learning model cannot. In other cases machine learning models can detect earlier than humans. This relies to a great extent on the nature of the specific conversation, as no conversations are equal. What is interesting is to find out more about the times humans detect earlier than machine learning models. It is desirable to better understand what knowledge forms the basis for this decision, and find out if cyber grooming detection systems based on a machine learning algorithm potentially can take advantage of knowledge extracted from human analysis.

1.4 Justification, Motivation and Benefits

Predatory adults are using the power of anonymity and the internet to their advantage in order to establish relationships with children and minors through cyber grooming. These relationships are further misused in order to perform inhuman acts to the children and minors, potentially harming them for life [2, 3].

Children are entitled to have a safe childhood, and parents should not be constantly worried about their children's presence online.

Live detection of cyber grooming contributes to detect potential dangerous conversations between children and adults pretending to be children. This detection makes it possible to stop the conversation from unfolding any longer. Further this contributes to avoid sexual abuse and malicious actions from taking place, protecting children online from getting their lives potentially destroyed.

In addition to first of all protect children online, cyber grooming detection also secondly saves families from a lot of sorrow and frustration. Also, it saves society for both money and resources as professional help will not be needed to handle harm and trauma caused by situations which degenerates from cyber grooming. Last but not least, cyber grooming detection makes it possible to provide information and documentation for legal authorities to take legal actions against predatory adults, which can hopefully stop the person from performing such actions ever again. Widespread information about cyber grooming detection can also have a preventive effect on others, hopefully scaring them to not perform any such illegal actions at all.

1.5 Research Questions

This section will introduce the defined research question we want to answer throughout the master's thesis. The research question is also divided into some smaller sub-questions to be answered in order to better answer the main question.

Research question:

- Can a cyber grooming detection system based on a machine learning algorithm be improved utilizing knowledge extracted from human analysis?

In order to answer the research question, we have defined some smaller sub-questions which look at parts of the research:

- What features of predatory and non-predatory conversations do humans react to in order to evaluate them potentially predatory and non-predatory?
- What features from human analysis can be used to improve a cyber grooming detection system?

1.6 Planned Contributions

The contribution of this master thesis will be extended knowledge of features found from human analysis for detection of potentially predatory and non-predatory conversations. These features will be of defining nature, making them suitable to use for the purpose of cyber grooming detection and in cyber grooming detection systems. The features can potentially increase the detection speed and detection rate of predators in cyberspace. As a consequence of this improvement, it will become more safe for children and minors to be present online and the probability for potential sexual abuse or malicious actions will decrease.

Chapter 2

Background

This chapter provides an overview of state of the art literature related to the research question and sub-questions identified in section 1.5 of this report.

In order to get a good understanding of the thematic of this research project, it is necessary to break it down into a few key elements. These fundamentals will in this chapter be further explained to substantiate the understanding of cyber grooming detection.

2.1 Grooming

Grooming is the preparation process where a person, which we also will refer to as a predator, prepares the ground for sexual abuse of a child or minor, which we will refer to as a victim. Through this process the predator prepares the victim and its surroundings to facilitate the intended abuse or malicious act. The preparation consists of, but is not limited to, getting access to the victim, establishing a relationship, trust and confidence, and making sure the victim keeps the communication to it self in order to circumvent any others from discovering the intentions of the predator minimizing the risk of getting caught [5]. Throughout this multistep process, a variety of techniques and tactics are used by the predator to reach its goal of performing sexual abuse or malicious actions [6].

The legal aspects of grooming is somehow intricate as there is a fine line between what is defined as legal and illegal activity by law, which is an essential part in order to convict someone for doing something. What makes it even more complex is the fact that the law is different in different countries. This makes it harder to have one common definition of what is to be considered illegal world wide. In Norway, the Norwegian Criminal Law § 306 [7] defines it as a criminal offence to plan a meeting with a child with the intention of performing sexual abuse. In other countries than Norway, where the age of consent is 16 years old, the age of consent varies all the way from 11 to 21 years old [8]. Also the action of abusive behavior against a person under 16 years old is defined as a criminal offence in Norway [7]. Such legislation forms the basis for what needs to be detected in cases of grooming and cyber grooming for further analyzes by human moderators and law enforcement. But how can we with certainty claim that someone has the intention of performing sexual abuse or any other illegal activity?

It is obviously very hard to determine if someone has the intention of performing sexual abuse or other malicious actions with a child or minor in case of a meeting, or if the intentions are pure harmless. It is simply impossible to read someones mind. As long as no such thing as sexual act or similar, or any other direct indicators of it is presented throughout the communication, it is not illegal to be friends, hang out, and be genuinely nice. It is in many situations, however, considered to be strange and suspect for an adult person to initiate a friendship with a child or minor, especially if it is random and they do not have any legitimate reason to be friends [5, 9]. There are cases where people are defined as adults by their years of age, but are having the mental age of a child [10]. This is not the most common situations, but a vital point to consider in case of an adult with this mental state approaches a child to become friends. It could be stated that laws around the world are maybe not specific enough as there is room for interpretation, and further that the law is maybe not adequately guarding the children and minors of the world society [5, 9]. Luckily it is a thematic with increasing attention and is described as a priority by the EU. The EU are working constantly to improve the safety of children. For the period 2020-2025 they will work on creating a robust legal framework, step up the law enforcement response, and gather the many actors working for child protection and support in order to coordinate the work for the best result in a combined force [11].

Sexual abuse during childhood is very serious and often results in scars for life for the victim, both physically and psychologically. Research shows that victims of sexual abuse during childhood to a greater extent suffers from other similar happenings, like domestic violence and subsequent rapes or sexual assaults later in life. For the victims experiencing sexual abuse during their childhood, life could become a living nightmare. Even if they are not exposed to subsequent physical abuse or actions, psychological lifetime traumas could be as bad or even worse, resulting in a totally destroyed quality of life. Such psychological consequences could be post-traumatic stress disorder symptoms (PTSD); aversion from social happenings, depression, anxiety, learning and behavioral troubles, suicide attempt, abuse of alcohol and stronger drugs, and other more or less serious ailments [2].

2.1.1 Cyber Grooming

With the constant technological development in the society, humans are introduced to a lot of new opportunities. One such opportunity that have totally changed how the world works over the last decades, is the introduction of the internet. The internet provides the opportunity to easily communicate across the world and a message could be sent to the other side of the world in the blink of an eye. The internet has become a place for all sorts of things, including socialising, and it attract all sorts of people. Before the internet, we had to go out to meet new people, which we still can, but now we can also meet new people through various websites, online forums and a wide variety of social platforms. The world has in some extent become a re-shaped place compared to how we knew it before the 90's and it is more connected than ever. Getting to know new people is great and allows for getting to know people we probably would never have met in the real world. This is becoming more and more normal, as internet connected devices are becoming a bigger and bigger part of our daily lives, both at work and private [1]. This does not only apply for adults, as also children are becoming more and more exposed to connected devices like tablets, smartphones, computers, and gaming consoles at a very early age. With the use of all sorts of connected devices, the use of internet and online communication comes as a natural consequence and research shows that internet use by children mostly increases year by year. This is to be considered as a natural trend, as e.g. more and more learning activities are available through internet connected devices. School work is the number one of common things children uses the internet for, which makes internet a necessity in order to progress in school, and the also a very natural part of the daily life [12]. Throughout 2020 and 2021 we have also been witnessing the pandemic of Covid-19, which to an even greater extent has forced children online in order to be able to keep in touch with their friends and attend school classes.

Cyber grooming builds on the same fundamentals as ordinary grooming [5], but the important difference between the two is where it happens and how it happens. As ordinary grooming takes place offline in the real world, cyber grooming on the other hand takes place online in the cyberspace. This by utilizing the communication possibilities provided by the internet, as well as all of the other advantages and disadvantages provided. As the internet has become more and more common, piles of different communication platforms have emerged. We are now allowed to communicated through a wide variety of different online communication platforms for all sorts of purposes. We use Facebook to keep in touch with friends and family, Twitter to share knowledge, opinions and all sorts of things, Skype to call or chat, Messenger to chat, Tinder to date, and loads of other platforms for the same and different purposes [1]. With the steady increase in the presence of children online [12] also predators follow and adapt to all sorts of new opportunities online to be able to reach out to their victims and potential new victims, and they are really creative utilizing the online possibilities to the fullest [1].

In addition to the fact that cyber grooming happens online, it differs from offline grooming in especially one significant way. One of the major powers of the internet is the ability to be anonymous, which is one of the most important tools for the predators in their approach to children or minors online, as it makes it possible to hide their real identity. By creating false user profiles, the predators can pretend to be someone else than they really are. This is done by using a fake name, fake age, fake gender, fake profile picture, and whatever fake information needed in order to create a fake profile and appear to be another person. Some places it is not even necessary to enter any information in order to create a profile. Only a pseudonym could be enough, and the predator can then give away fake information when needed throughout the communication with the victim, in order to appear to be more attractive to the victim [13]. Throughout the communication the predator tries to gain as much personal information about the victim as possible in order to get a good overview of the victim and to be able to adapt the behavior to suit the victim and its needs. This is further used by the predator to gain trust and advantage over the victim [13, 14].

As the internet has emerged and people have found new ways of doing old things, as well as found new things to do, legislation has become outdated as it has not been created to take into consideration the new opportunities and ways of doing things on the internet. There are examples of events where old and outdated legislation has made it impossible to convict someone for doing something that we clearly consider to be illegal. This because it is not defined by any legislation, and it has therefore been impossible to convict someone for it, as no legislation has covered it. Luckily the general awareness of cyber grooming, and generally happenings in cyberspace, has increased. This has led to an increased focus on adapting, improving, and keeping the legislation up to date [5]. As with offline grooming, the work of the EU is also as much to prevent cyber grooming [11]. Such awareness is essential in order to be able to fight the predators and avoid cyber grooming potentially resulting in child abuse and trauma. It is important to keep a proactive approach and try to be in front of their next moves, so legal actions can be performed immediately. This is important in order to get the predators, but also as a preventive measure to hopefully scare others from performing illegal actions and becoming predator [9, 11].

2.2 Machine Learning

Machine learning is a vital part of many systems for cyber grooming detection, so it is essential to understand the fundamentals of this topic as well as more in-depth of certain sub-topics of it utilized in cyber grooming detection.

As humans we learn through experience and knowledge. Some of which is passed on to us from older generations, and some are new discoveries often created by utilizing previous experience and knowledge in order to get a new understanding. From observation we collect a lot of data which we bring on further to analyzing and utilizing in order to create predictions or new understandings. In such a way, we always use data to learn and gain new experience and knowledge. This is also the basis for all scientific work, it is based on learning from different types of data using different types of focus and approaches. Machine learning is nothing different and builds upon the same principles as human and scientific learning [15]. But humans are limited in some ways when it comes to processing capacity as the amount of data we need to process increases. This is a problem that is solved with the invention and development of computers and computer software, which allows for processing of much bigger sets of data very much faster than humans are able to do. Machine learning is a way of utilizing computers as a method for processing huge amounts of data and information. By utilizing different types of algorithms, machine learning is used to increased performance or make good predictions for the future. The data could be all sorts of data collected on a topic and put together for analysis (for the purpose of cyber grooming detection in chat, such data means huge amounts of text data collected from chat logs from different chat services available online). The data used for learning is often referred to as training sets. Based on the learning method of choice, the training sets are either labeled by humans or structured in some other way through environment interaction. The size and quality of the training sets are of great importance, as they are crucial for the machine learning model to be trained as good as possible and further be able to perform as good as possible when used for its purpose. An example to state the importance of size and quality of training sets is if you want to learn someone what a horse looks like. The person does not know what a horse looks like from before and you are only allowed to use pictures. The more pictures you can show of horses, the easier it will be for this person to tell if he sees a horse at a later time in life. So you would for example like to show several different types of horse breeds with different colors, and pictures from different angles and distances, and you would also prefer to have pictures of as high quality as possible. Another important factor you want to make sure is that the training set of horse pictures are of good quality, i.e. you want only pictures of horses, not one or more pictures of cows, pigs, or other animals or things to confuse the person you teach. This is also the same principals that is used when training a machine learning model [16].

Machine learning deals with different types of learning problems, which have different learning methods. Supervised and unsupervised learning are two learning methods out of several other. Supervised learning is when the training sets are labeled and is often used to handle classification, regression, and ranking problems. A labeled training set is a set of data where e.g. 100 out of 200 pictures are labeled as "dog" and the reminding 100 are labeled as "not dog". In other words, labeled training sets are prepared in advance so the label can tell the machine learning algorithm what is in the picture, and the model further can find patterns and features of all pictures labeled the same. This is to create rules for what the model should recognize as "dog" and "not dog". Unsupervised on the other hand is when the training sets are not labeled and is often used for clustering and dimensionality reduction problems. For unsupervised learning without labels, the machine learning algorithm will have to find and group pictures where similar features and patterns are found in order to try creating rules for the model [16].

From the learning performed by the algorithm on training sets, the machine learning model creates equation systems, rules, relations, functions, probability distributions, and other representations of knowledge. After training a machine learning model it can be used to perform the task it was created and trained for, like detecting if pictures contains horses or not [17].

The three next subsections will explain two important types of algorithms; classification and regression, in addition to data preprocessing. These terms are important as they are frequently used in the area of cyber grooming detection, and in combination with natural language processing, which will be further explained in next section.

2.2.1 Classification

Classification problems are the most frequently used ones in machine learning [17]. As explained, classification problems are trained on supervised methods where the training sets are labeled. This means the data sets are structured due to its content. For example in a labeled training set for classification of animals, the pictures will be structured and organized based on what animal is in the picture. This is what we refer to as classification, as each animal will represent its own class containing pictures of the defined animal [16]. The classification problem is then to determine the exact class for a new, unknown and unlabeled picture (object) from a total number of possible classes [16, 17]. In order to perform good classification of objects, it is essential to have a sufficient number of attributes (features, properties) which are independent observable variables, which are either discrete or continuous. This makes the classes more defined and it becomes easier for the machine learning model to create good classifiers for the different classes. Each class consists of dependent unobservable discrete variables with value based on the respective independent variables. Good classifiers are essential for a machine learning model to be able to predict what class a totally unknown object belongs to based on its attributes, since these are the data points the model has to base its decisions on. Weak classification will result in a model not working very well classifying objects wrong, causing false negatives (FN) and false positives (FP). False Positive, also called Type I Error, is when a model claims something is true, when it is in fact not true. For example a cyber grooming detection model can state a conversation is predatory, when it actually is non-predatory. False Negative, also called Type II Error, is when a model claims something is not true, when it is in fact true. For example if the same cyber grooming detection model claims a conversation is non-predatory, when it actually is predatory. In order for models to be able to perform classification, the classifier needs mapping between the attribute space and the class space. Such mapping can be done in many ways, and is performed by a discrete function described by the classifier [17].

There are several common classifiers for classification of data, like decision trees, decision rules, Naïve Bayesian classifiers, nearest neighbor classifiers, logistic regression, support vector machines and artificial neutral networks [17].

2.2.2 Regression

Regression problems also starts with a set of objects with the associated independent observable variables which could be continuous or discrete; attributes (features, properties) [17]. For regression the dependent variable is continuous (not discrete) with a value based on a function of independent variables [15, 17].

In a classification model we get an output classifying if an object belongs to a certain class or not, which is a question of yes or no. Regression on the other hand differs from classification as it outputs a predicted value for the dependent unobservable continuous variable for the specific object. This function could either be learnt from problems solved earlier or given beforehand. [15–17]. For a regression based learning algorithm, the mission of the algorithm is to decide a continuous function by learning from training sets of data [17].

For regression based machine learning there are several common regressors like linear regression, regression trees, locally weighted regression multi-layered feedforward neural networks for regression, and support vector machines for regression [17].

2.2.3 Data Preprocessing

In order to get the best possible performance out of the machine learning model, data preprocessing is essential for the training set to be as good as possible. Preprocessing is the preparation of the data of interest in order to arrange it in a way that allows for getting the most out of it through the training. There are lots of ways to prepare data and the preparation should be relative to the intended use. Cleaning, selection, transformation and feature extraction are examples of some actions that can be performed to prepare the data [17].

2.3 Natural Language Processing (NLP)

Natural language processing (NLP) is the process of making human language readable for computers and is based on several other sciences, like algorithms, linguistics, logic, and statistics [18, 19]. The human language is easy, but yet so complex and complicated. It is our most important tool for sharing information and knowledge from one person to another. This has been done for thousands of years, from generation to generation. But when you introduce computers to the equation, things are getting complicated. Natural languages are not made for being interpreted into a finite set of mathematical operations, and computers are created for handling 1's and 0's humanized through different programming languages, not process natural languages. With the use of NLP, computers are capable to first of all read the language, but further also to derive meaningful information that could be used for different purposes [19]. In this project we aim to derive valuable knowledge from human analysis of conversations in order to find defining features that potentially can be used to improve cyber grooming detection systems utilizing natural language processing technology. By adding knowledge from human analysis, hopefully systems can improve the natural language processing capabilities of their models, and further the total functioning of the models.

Natural language differs from computer languages in especially one significant way, namely that they can be ambiguous and have several meanings, e.g. through the use of sarcasm. This is essential to take into consideration when working with NLP in order to interpret the correct meaning [18, 20].

For the computer to be able to do anything with natural language input data, it needs to be extracted into structured numerical data as vectors by utilizing linear algebra. From vectors it is possible for computers to perform mathematical operations and utilize the data for machine learning. The possibility of storing "meaning" of text also comes in handy instead of just characters and words, which further with semantic analysis helps interpret the ambiguity of natural languages [19].

2.3.1 Bag of Words

Bag of Words (BoW) is a method where the occurrence of every word in a text is counted and put into a dictionary, or "a bag of words", without considering the structure or order of which the words occurs. Only the word count is considered, nothing else. It is also common to use an already existing dictionary, created from multiple other texts. For this dictionary, the text in question is turned into a sparse vector with the same length as the size of the dictionary. From start, each entry of the vector is defined as 0. When the index points to a word occurring in the text, also present in the dictionary, the value of the vector is updated. The value could then be either binary 1, meaning the word occurs at least once throughout the text, or it could be an integer value indicating occurrence of n times throughout the text (Term Frequency (TF)). It is quite effective using the Bag of Words method for classification of text and it is commonly used in NLP [18].

By utilizing this technique it is possible to detect documents that are similar due to what words are used, and then be able to extract meaning based on the content of several documents where the text and content is of similar art [21].

2.3.2 TF-IDF

TF-IDF is commonly used technique in NLP [19, 22] and stands for Term Frequency - Inverse Document Frequency. It is composed of two concepts, term frequency (TF), and inverse document frequency (IDF). Term Frequency is the number of times each and every word occurs in a single document. Document Frequency (DF) is the total number of documents out of a collection, where a term t occurs. Inverse Document Frequency is the word occurrence for each word divided by the total number of documents the current word occurs in. By performing such calculations it is then possible to say something about the relevance of words, and further documents of a corpus [19].

In order to avoid bias of longer documents, term frequency normalized for that matter is given below where the numerator $n_{i,j}$ represents the total number of occurrences of term t_i in document d_j, and the total number of occurrences of all terms for all documents d_j represented by the denominator [21]:

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \tag{2.1}$$

Inverse document frequency can be written the following way, where the numerator N represents the number of documents over the denominator representing the number of documents containing the term t_i [21]:

$$idf_i = log \frac{N}{\mid \{j : t_i \in d_j\} \mid} \tag{2.2}$$

To generate TF-IDF, simply find the product of term frequency and inverse document frequency [21]:

$$tf\ idf_{i,j} = tf_{i,j} * idf_i \tag{2.3}$$

2.3.3 Word Embedding

BoW and TF-IDF have long sparse vectors. The size of the vectors relates to the size of the dictionary in use, which can easily be of lengths around 20.000-30.000 words. Word embedding typically have vectors of a value k between 100 and 500 in length. These are much smaller than the vectors used by BoW and TD-IDF, and are not sparse, which makes them easier to use for calculations. Word embedding is the representation of words using k-dimensional vectors of real numbers [18]. By using such mapping it allows for similar representation of words with similar meaning [23].

Word2Vec

One application of word embedding often utilized in NLP is Word2Vec by Mikolov et al. [24]. Word2Vec is a distributed representation due to the semantics captured for the word by the activation pattern of the full representation vector. By taking advantage of the context of a defined word, Word2Vec is able to learn its semantics. It also looks at surrounding words to the word of attention in order to learn the numerical representation of it [25].

FastText

FastText was introduced in 2016 by Facebook, inspired by Word2Vec. As Word2Vec gives individual words to the neural network, FastText creates n-grams (sub-words) by breaking down the words before they are passed on to the neural network. The word embedding vector representing the word then contains the total number of n-grams for the given word. This way of doing it allows for a better representation of rare words because it is more likely that some of their n-grams are present in other words. Also, FastText can find representation of new or misspelled words, which differs from BoW, TF-IDF and Word2Vec [19, 23].

2.4 Statistics

Statistics is an important part of machine learning and NLP, as they are built on utilization of different statistical concepts. It is also essential for measuring different types of performance and results. For this reason, statistics is essential in order to be able to measure how well machine learning models for cyber grooming detection works, and compare them to the performance measures of other models, to see if improvements will do any good or bad.

Statistical measures of performance are used in several researches on cyber grooming detection, like e.g. [3, 13, 26].

2.4.1 Precision, Recall and F-score

Precision and recall are two nuanced metrics for evaluation of classification models [27]. The two are frequently used in different research on cyber grooming detection [3, 13, 26], making it easier to compare across different models and researches.

Results from classifiers are displayed in a confusion matrix showing the total of correct and incorrect predictions sorted by response. Table 2.1 illustrates the confusion matrix. True Positive (TP) is when e.g. a conversation is stated to be predatory, when it indeed is predatory. True Negative (TN) is when e.g. a conversation is stated not to be predatory, when it indeed is not predatory. False Positive (FP) is when e.g. a conversation is stated to be predatory, when it in fact is not predatory. False Negative (FN) is when e.g. a conversation is stated not to be predatory, when it in fact is predatory. [27].

	Predicted Response	
True Response	True Positive (TP)	False Negative (FN)
	False Positive (FP)	True Negative (TN)

Table 2.1: Confusion matrix

Pure accuracy is the fraction of correct predictions [27]:

$$Accuracy = \frac{\sum TP + \sum TN}{\sum TP + \sum FP + \sum TN + \sum FN} \tag{2.4}$$

Precision is the measure of accuracy of predicted positive outcome, i.e. the fraction of actual relevant elements from the total set of items predicted as relevant by the model. In other words, how sure can we be that the stated relevant elements are actual relevant. It is defined the following way [27, 28]:

$$Precision = \frac{\sum TP}{\sum TP + \sum FP} \tag{2.5}$$

Recall is the measure of how strong the model is, and describes a models sensitivity, i.e. what is the probability that a relevant element is actually detected by the model. It is defined the following way [27, 28]:

$$Recall = \frac{\sum TP}{\sum TP + \sum FN} \tag{2.6}$$

Precision and recall are not perfect as standalone measures as they can be tricked into giving perfect or misleading answers. This is due to the nature of how they are mathematically composed. For example, if the model always outputs that an element is relevant, then the recall will be 1 because there are no FN, without having any significant contribution. For precision on the other hand, if a model always outputs that an element is relevant, then the precision will be 1 because there are no FP, also in this case without having any significant contribution at all [28]. In order to deal with this problem it is essential to balance the two measures together with a harmonic mean. This balancing is what we call F-score [29]:

$$F_\beta = (1+\beta^2)\frac{Precision * Recall}{(\beta^2 * Precision) + Recall} \tag{2.7}$$

The harmonic mean is what we call the F_1-score, i.e. $\beta = 1$. By increasing the value of β, recall is favoured, and by lowering the value beneath 1, precision is favoured. Adjusting β also allows for avoidance of undetected relevant elements or avoidance of false accusations [28].

Accurate detection is essential, but what is at least as important is to detect cyber grooming as early as possible. It is desirable to detect such activity after as few messages exchanged as possible. Cyber grooming is challenging in many ways because of how it unfolds. Each conversation is different, and it can be everything from grooming from the very beginning of the conversation, and all the way to harmless for several years before the conversation stands out as grooming. There are currently no good way for measuring performance of speed for cyber grooming detection, i.e. how few messages is needed for detection. This is intricate and challenging.

2.5 Cyber Grooming Detection

The scope of this project is towards detecting cyber grooming in one-on-one chats from various chat platforms online. A lot of research has been performed in the area of interest, and it is still a very relevant and hot topic for research. However, most work on the topic has been based on complete conversations in hindsight, which makes it too late in order to be able to prevent sexual abuse or malicious actions from happening. Newer research have started trying to detect predatory behavior in real time. By analyzing message by message, detection models based on machine learning makes it possible to detect a potentially predatory conversation, which can be further analyzed by a human moderator. If it is found to be predatory by the human moderator, it can be reported to law enforcement's for further handling of the situation in order to avoid sexual abuse or malicious actions from happening. In this section we will take a closer look at some of this research and how detection is performed.

As for all machine learning models, it is essential to use quality training data to get the best performance and results possible. Most research used predatory conversation chat data from Perverted Justice [30] for training their models. The PAN-2012 competition [31] contained data from [30] as well as non-predatory conversations from other chats. The data from [30] provides complete chat logs as transcripts of known predatory conversations proven by conviction.

In 2012 the International Sexual Predatory Identification Competition was held at PAN. The competition presented the participants with two problems. Problem 1 was to identify as many predators as possible from provided data sets containing chat logs with both normal conversations and proven predatory conversations. Problem 2 was to identify the prominent predatory lines from the provided conversations. Several teams participated in the contest and provided solutions to the problems. For problem 1 different techniques for pre-filtering the data was used followed by a two stage classifier. In some cases the first stage classifier was used to determine whether or not conversations were predatory (true positive) or non-predatory (false negative). This step was necessary to filter out false negatives as the datasets were design to be unbalanced (heavily weighted with false negatives) in order to reflect a scenario as realistic as possible. The second stage classifier separated the victim and predator in conversations that turned out to be suspicious [31, 32].

Throughout most of the submissions for the contest, the features could be divided into two main categories: lexical and behavioral features. Lexical features are features extracted from the raw text from conversations. Behavioral features are features concerning the actions of users withing conversations. Further in the classification step several methods were utilized, like Neural Network classifier, decision trees, Naïve Bayes and more. The mostly used method was Support Vector Machines (SVM), but in some cases other classifiers, like Neural Network classifier, outperformed the SVM [31].

For problem 2, no training data were provided, making it more challenging to test the participants. Most solutions utilized their findings from problem 1 to find all predatory lines of conversation. Further this was filtered through a dictionary of perverted terms or using particular score from e.g. TF-IDF weighting [31].

Valuable knowledge about cyber grooming detection was derived from all the participants and their submissions to the competition. As for features, Inches et al. stated that both lexical and behavioral are of great relevance in such context and both should be used. Also the use of pre-filtering is essential to remove conversations of no interest. Regarding method for detection of specific lines, several methods showed to provide good results, hence there is no single method best suited for detection of cyber grooming and predators [31].

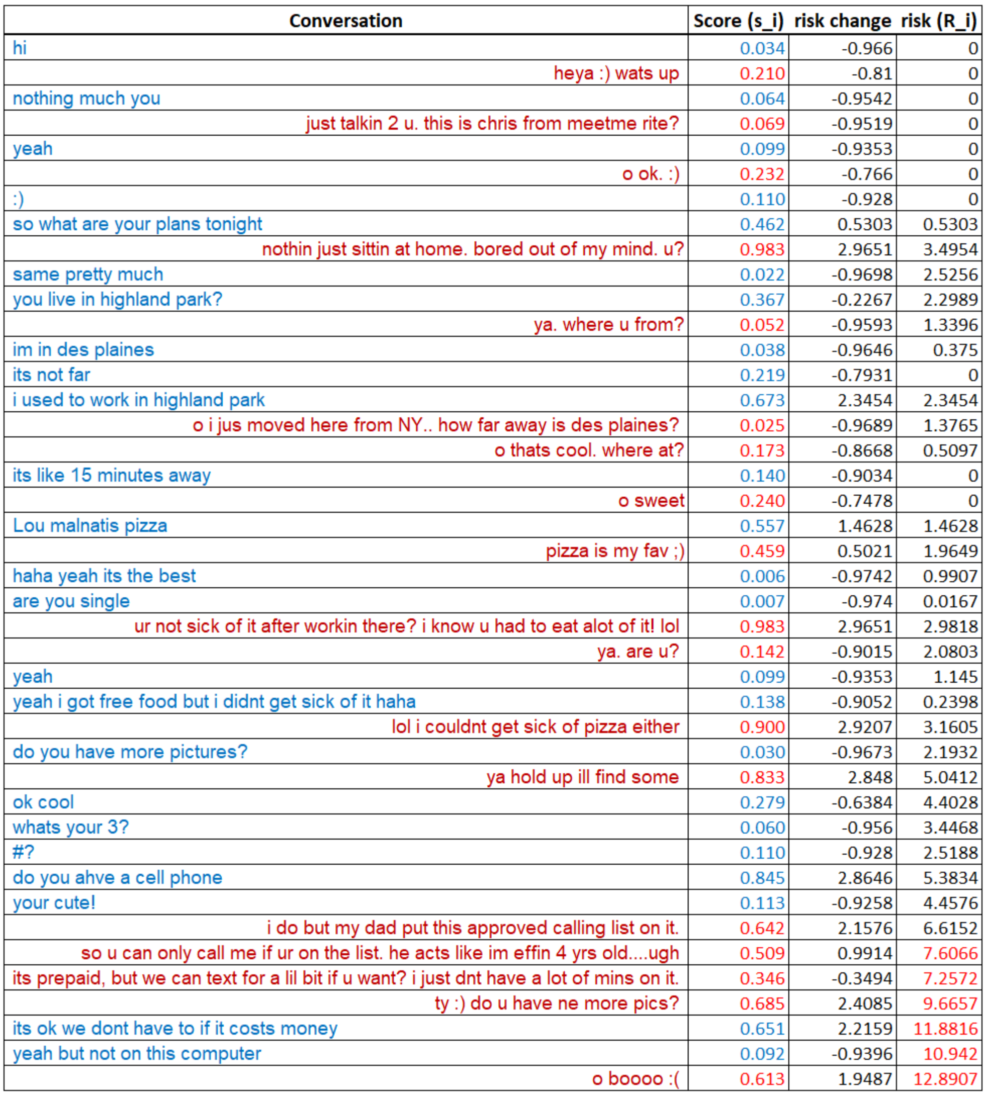
Michalopoulos et al. [4] presented a system called Grooming Attack Recognition System (GARS) to perform real-time identification, assessment and control of cyber grooming attacks in order to increase the online security of children. The system utilizes multiple methods to generate a total risk value which is continuous updated based on chat conversations. When the risk level reaches a certain threshold, a warning is issued instantly to e.g. the child's parents, and is also displayed to the child. In order to evaluate the risk, the system uses document classification, personality recognition, user history and exposure time [4].

One key element of cyber grooming detection is to detect if an adult is posing as a child while communicating with an actual child. Ashcroft et al. [13] studied whether or not it is possible to determine if the person writing is a child or adult based on writing style, and further to evaluate if the detected child is an actual child, or an adult person impersonating a child. Through this work they found it possible to get good accuracy when distinguishing between children and adults as long as the text language is formal. With more informal writing (e.g. blog text and chat logs), the separation is harder, but in almost all cases they were able to distinguish children and adults impersonating children from chat log data. [13].

Bours and Kulsrud tried to solve the problems of PAN-2012, and by comparing their results to the top 10 contestants from [31], they produced results performing good. On the first phase they used the Conversation-Based Detection (CBD) method with SVM on the TF-IDF features from the complete conversations. On the second phase they utilized Ridge classifier on TF-IDF features on the individual parts of the conversation. Their result, compared to the PAN-2012 competition [31], would have placed them 3rd on precision (0.891) and $F_{0.5}$-score (0.887), 2nd on recall (0.870) and F_2-score (0.874), and 1st on F_1-score (0.880) [3].

Through their research, Bours and Kulsrud found that models for Author-Based detection combined with Neural Network classifier worked good, and also that 2-phase CBD method in combination with Ridge classifier or Naïve Bayes classifier resulted in good performance. The research showed that it is possible to detect cyber grooming live at a quite early stage, which is essential in order to be able to prevent sexual abuse or other unwanted situations [3].

As most research bases the detection on complete conversations, Bours and Kulsrud saw the need for continuous live detection of predatory conversations as quick as possible, and in 2019 they proposed a solution to the problem. Because time is essential in situations of cyber grooming, they focused on developing a system able to detect a predatory conversation as quickly as possible in order to further take the necessary steps to avoid sexual abuse. Their system, called AiBA (Author input Behavioural Analysis), analyzes each and every message sent between two parties, by utilizing machine learning models. They calculate a risk value R_i for the first i messages in a conversation and the risk value is updated after every new message. Initially, a conversation starts at risk level 0, i.e. $R_0 = 0$. Each message will be evaluated by a machine learning model and will receive a score between 0 (innocent) and 1 (dangerous). If the score of the machine learn-



Conversation	Score (s_i)	risk change	risk (R_i)
hi	0.034	-0.966	0
heya :) wats up	0.210	-0.81	0
nothing much you	0.064	-0.9542	0
just talkin 2 u. this is chris from meetme rite?	0.069	-0.9519	0
yeah	0.099	-0.9353	0
o ok. :)	0.232	-0.766	0
:)	0.110	-0.928	0
so what are your plans tonight	0.462	0.5303	0.5303
nothin just sittin at home. bored out of my mind. u?	0.983	2.9651	3.4954
same pretty much	0.022	-0.9698	2.5256
you live in highland park?	0.367	-0.2267	2.2989
ya. where u from?	0.052	-0.9593	1.3396
im in des plaines	0.038	-0.9646	0.375
its not far	0.219	-0.7931	0
i used to work in highland park	0.673	2.3454	2.3454
o i jus moved here from NY.. how far away is des plaines?	0.025	-0.9689	1.3765
o thats cool. where at?	0.173	-0.8668	0.5097
its like 15 minutes away	0.140	-0.9034	0
o sweet	0.240	-0.7478	0
Lou malnatis pizza	0.557	1.4628	1.4628
pizza is my fav ;)	0.459	0.5021	1.9649
haha yeah its the best	0.006	-0.9742	0.9907
are you single	0.007	-0.974	0.0167
ur not sick of it after workin there? i know u had to eat alot of it! lol	0.983	2.9651	2.9818
ya. are u?	0.142	-0.9015	2.0803
yeah	0.099	-0.9353	1.145
yeah i got free food but i didnt get sick of it haha	0.138	-0.9052	0.2398
lol i couldnt get sick of pizza either	0.900	2.9207	3.1605
do you have more pictures?	0.030	-0.9673	2.1932
ya hold up ill find some	0.833	2.848	5.0412
ok cool	0.279	-0.6384	4.4028
whats your 3?	0.060	-0.956	3.4468
#?	0.110	-0.928	2.5188
do you ahve a cell phone	0.845	2.8646	5.3834
your cute!	0.113	-0.9258	4.4576
i do but my dad put this approved calling list on it.	0.642	2.1576	6.6152
so u can only call me if ur on the list. he acts like im effin 4 yrs old....ugh	0.509	0.9914	7.6066
its prepaid, but we can text for a lil bit if u want? i just dnt have a lot of mins on it.	0.346	-0.3494	7.2572
ty :) do u have ne more pics?	0.685	2.4085	9.6657
its ok we dont have to if it costs money	0.651	2.2159	11.8816
yeah but not on this computer	0.092	-0.9396	10.942
o boooo :(	0.613	1.9487	12.8907

Figure 2.1: Message score and risk development throughout a conversation using AiBA

ing model of message i is denoted s_i, then the risk level is updated as a function of the old risk level and the score of the new message. In other words $R_i = f(R_{i-1}, s_i)$. The risk will increase in case of a dangerous message and decrease with an innocent message, but the risk level will never drop below 0. Figure 2.1 shows how the risk changes throughout a conversation. The maximum increase or decrease of the risk do not need be the same. When the total risk grows above a certain threshold, a human moderator is warned to further evaluate if the conversation is predatory or not, and if needed reported to law enforcement. In the example given in figure 2.1, the threshold is defined at 7.0. Figure 2.2 illustrates how the total risk changes throughout the conversation. In a slightly different analysis, they managed to detect predatory conversations after 40 messages on average, while the full conversations were on average over 3000 messages long [3]. This approach to live detection of cyber grooming using total risk score is also very similar to the GARS system presented by Michalopoulos et al. [4].

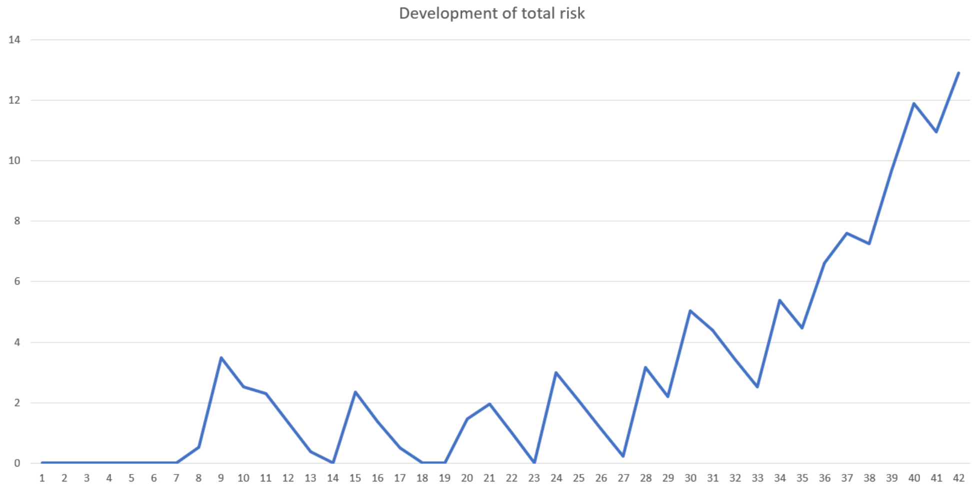


Figure 2.2: Total risk development graph in AiBA

Chapter 3

Data

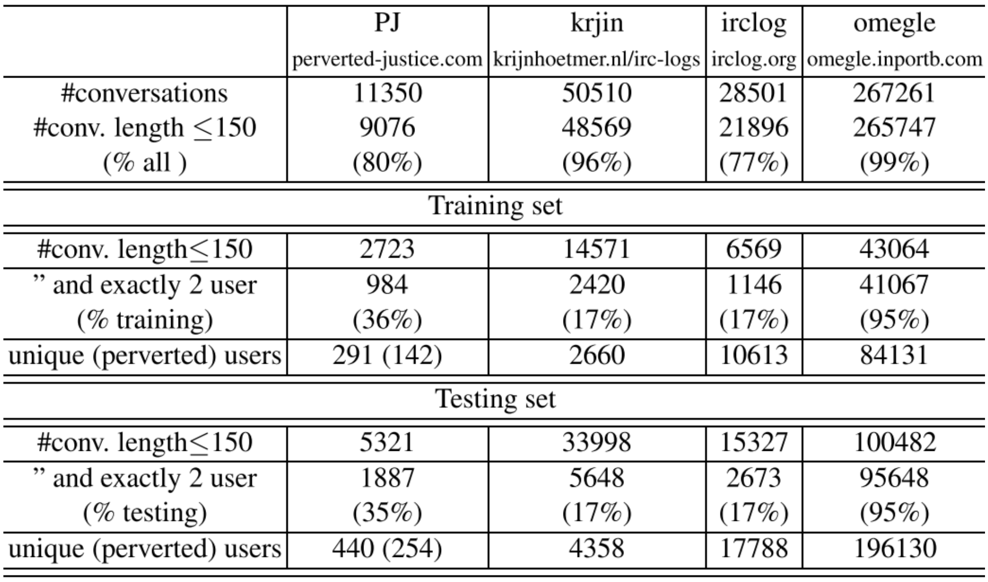
This chapter explains the data used in this project and where it comes from. It also explains the experiment conducted for getting human evaluations of conversations, and how the gathered data was prepared and combined with data from the other datasets.

3.1 PAN-2012

The data used throughout the work for this thesis originates from the PAN-2012 dataset, which is explained in detail in [31]. This dataset was used for the International Sexual Predator Identification Competition at PAN-2012. It contains real, historical chat conversations from chats gathered from different chat providers [31].

The PAN-2012 dataset was collected and created to be a highly valuable dataset to be utilized within different fields of research as a common point of reference. This allows researchers to compare the results of their different research approaches to each other in order to learn more about what performs good and what performs not so good. The dataset contains a large number of conversations, and aims to be as realistic as possible in comparison to the real world. This means the vast majority of the conversations in the dataset are non-predatory, and the predatory ones constitutes less than 4% of the total number of conversations in the original dataset. In order to be as realistic as possible, the dataset also contains different types of properties. The number of false positives (conversations which are sexual or within the same area as predatory conversations often could be) is large, the number of false negatives (normal conversations within many various subjects) is large, and the number of true positives (potentially predatory conversations) is low [31].

The true positives were collected from Perverted Justice (PJ) [30]. The conversations provided by Perverted Justice are conversations containing convicted sexual predators chatting with volunteers posing as underage teenagers [31]. The fact that the true positives contains real predators is very valuable, as it allows the



	PJ	krjin	irclog	omegle
	perverted-justice.com	krijnhoetmer.nl/irc-logs	irclog.org	omegle.inportb.com
#conversations	11350	50510	28501	267261
#conv. length ≤150	9076	48569	21896	265747
(% all)	(80%)	(96%)	(77%)	(99%)
Training set				
#conv. length≤150	2723	14571	6569	43064
” and exactly 2 user	984	2420	1146	41067
(% training)	(36%)	(17%)	(17%)	(95%)
unique (perverted) users	291 (142)	2660	10613	84131
Testing set				
#conv. length≤150	5321	33998	15327	100482
” and exactly 2 user	1887	5648	2673	95648
(% testing)	(35%)	(17%)	(17%)	(95%)
unique (perverted) users	440 (254)	4358	17788	196130

Figure 3.1: Properties of the original PAN-2012 dataset collected by [31]

experiment and analysis to consider real life behavior obtained from real predators. In such a way, we can be more sure of the patterns and features extracted.

In order to make the conversations in the dataset comparable, conversations where there was a pause of 25 minutes or more between messages exchanged were split. This means that one conversation from real life could potentially be represented as several separate conversations with unique conversation IDs in the dataset. What is important to note, is even though conversations were cut into multiple conversations, the IDs of the different chatters of the conversations remains the same. Also, conversations containing more than 150 messages exchanged were excluded from the dataset [31].

From the original PAN-2012 dataset, a new dataset was created by Bours and Kulsrud for their work with [3]. This dataset contains a selection of 32063 conversations from the original dataset. Conversations involving more than two chatters were removed, and so were also conversations involving only one chatter. After this only one-on-one conversations were left. The original PAN-2012 dataset also contains some conversations without any real content which were removed, where one of the chatters only kept repeating the same text over and over again. The PAN-2012 dataset also contains the ground truth data in the form of a list of predator IDs. This can be used to determine if a conversation is predatory or non-predatory.

3.1.1 Hybrid

From the dataset created by Bours and Kulsrud [3], a summer intern performed manual evaluations of 4084 randomly presented conversations to get a human's perspective on conversations. These 4084 conversations out of the 32063 constituted the initial basis for the experiment of this thesis. From these 4084, 2000 conversations were extracted into a new dataset which constituted the dataset used for the experiment performed for this thesis. As the experiment required participation from volunteers, the dataset had to be of a manageable size and not to comprehensive. The contribution required from the participants was quite comprehensive and required more from them than some ordinary survey typically will, both in terms of time and effort. Due to this fact, it was harder to get participants and to get them to do a sufficient number of evaluations, which further was the reason for creating a relatively small dataset compared to the ones it originates from. The smaller dataset aimed to get multiple evaluations of the same conversations, but this showed to be harder than initially thought. Even though the dataset was smaller, it was hard to get a sufficient number of evaluations from the participants.

The data used further for analysis consisted of the 4084 evaluated conversations plus the part of those which were evaluated through the experiment. This way, at least all conversations evaluated in the experiment were evaluated a minimum of two times. In combination, all of these evaluated conversations constituted the dataset used for this thesis analysis.

As the data originates from PAN-2012 and Bours and Kulsrud, the same list of author IDs proven predatory by Perverted Justice is to be used together with it. This allows to compare the evaluations to the ground truth. It is however worth noting that conversations classified as non-predatory from the other sources than Perverted Justice potentially can be predatory conversations. These conversations are in such case not proven predatory by court, but could potentially be predatory without anyone being convicted or caught for them.

3.1.2 Dataset Structure and Features

The hybrid dataset consists of a set of XML files. Each and every conversation of the dataset is represented as a single XML file, with a belonging file name defining the dataset name and a unique number for the file counting from 1 and upwards.

The XML files builds on the basic structure of XML as figure 3.2 shows an example of. The XML files are represented in a tree structure having a root (parent) and several branches, also called children. The root of each file is named "conversations". The following branch of conversations is called "conversation id" which also includes a unique ID for each file in order to be able to separate all XML files from each other. Within a conversation, there are one or several message lines representing each and every message sent back and forth throughout the conversation. This branch is named "message line" and also comes with a number describing when the message was sent in the conversation it belongs to, counting from 1 and upwards describing the chronological order. Each message also contains three branches holding metadata about each message. These are "author", "time" and "text". The first branch "author" holds a unique identifier for the party of the conversation that sent the message in question. Time tells what time of the day the message was sent, using a 24 hour time view. The last branch "text" holds the actual message that was sent in the message in question.

3.2 Data Collection from Experiment

The collection of data to be used for analysis in this thesis was conducted throughout an experiment. The experiment required volunteering participants for manually evaluating conversations from the 2000 conversation dataset.

The goal was to get as many evaluations as possible of the 2000 conversations. Preferably several evaluations of each conversation. This showed to be much more challenging than initially thought, as it was not easy to get volunteering participants in the first place, and it was even harder to get those who initially volunteered to actually do what they were supposed to do.

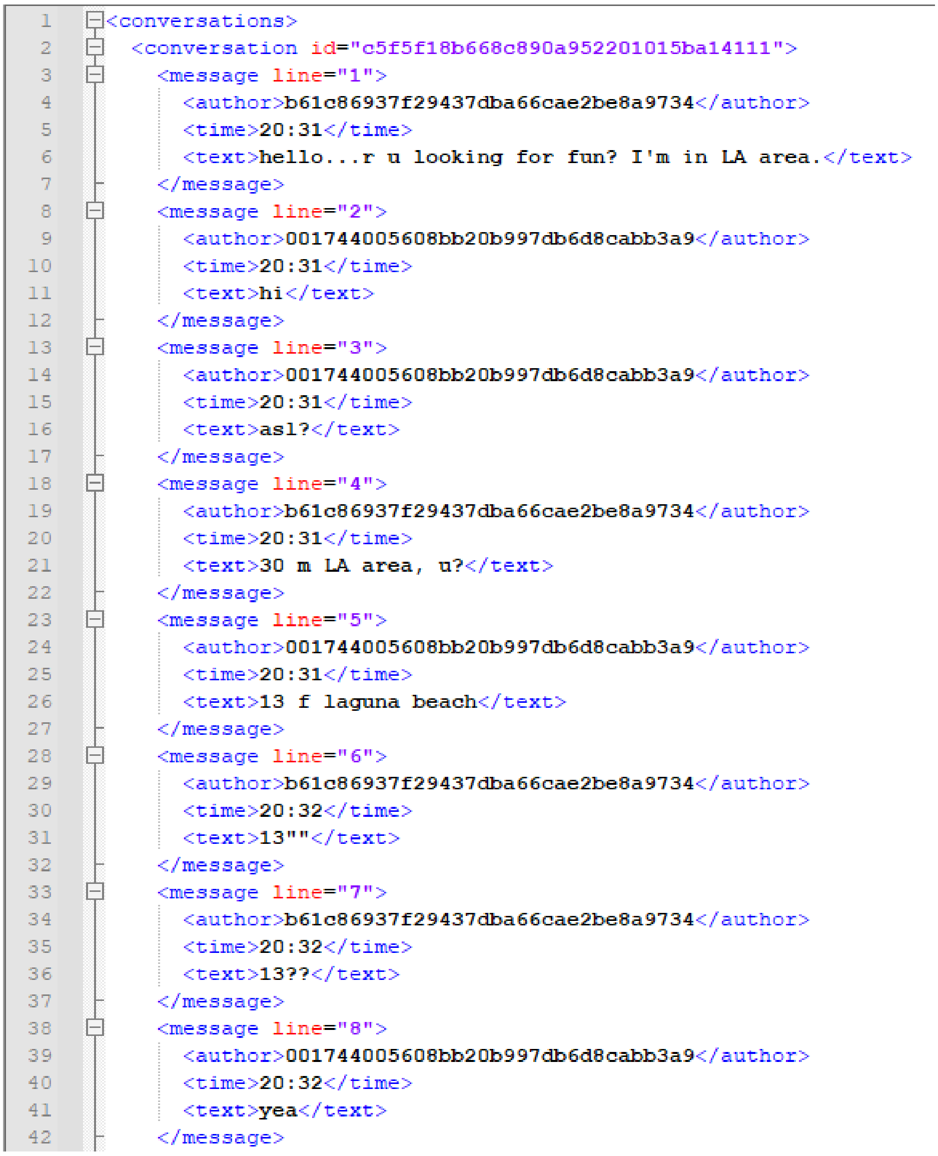
3.2.1 Participants

Selection of participants for an experiment is essential for the data collection to be as good and useful as possible. Since cyber grooming can be performed in many different ways, it is essential to cover as much ground as possible in regards of what is triggering human beings to evaluate a conversation to be potentially predatory. Older people have one way of viewing conversations based on their experience in life and understanding of the society today. Younger people, on the other hand, have another way of viewing conversations based on their experience through life so far, also being more used to chat as a communication platform. Also, it is possible that women will react differently than men. Based on this, the aim for this experiment was a wide variety of different participants, which hopefully would give valuable data for further use.

In order to avoid unnecessary feedback without substance, some limitations were set for the participation:

- The lower age limit for participation was set to 18 years old. The reason for this was that 18 is defined as the legal age in Norway and it was then not necessary to get approval from parents. Further, at the age of 18 people start to get more reflected due to experience in life, but still have youthful opinions and understanding. This is valuable in order to potentially get a better understanding for the meaning between the lines.
- The upper age limit was set to be 65 years old. Older people have grown up in another age without technology around, and are in general assumed not to be in possession of the desired knowledge and understanding needed for this study.

Gathering volunteers for participation showed to be quite much harder than initially thought. In total, we got 36 people to participate. First of all, it was challenging to convince people into participating after explaining the experiment.



```
<conversations>
  <conversation id="c5f5f18b668c890a952201015ba14111">
    <message line="1">
      <author>b61c86937f29437dba66cae2be8a9734</author>
      <time>20:31</time>
      <text>hello...r u looking for fun? I'm in LA area.</text>
    </message>
    <message line="2">
      <author>001744005608bb20b997db6d8cabb3a9</author>
      <time>20:31</time>
      <text>hi</text>
    </message>
    <message line="3">
      <author>001744005608bb20b997db6d8cabb3a9</author>
      <time>20:31</time>
      <text>asl?</text>
    </message>
    <message line="4">
      <author>b61c86937f29437dba66cae2be8a9734</author>
      <time>20:31</time>
      <text>30 m LA area, u?</text>
    </message>
    <message line="5">
      <author>001744005608bb20b997db6d8cabb3a9</author>
      <time>20:31</time>
      <text>13 f laguna beach</text>
    </message>
    <message line="6">
      <author>b61c86937f29437dba66cae2be8a9734</author>
      <time>20:32</time>
      <text>13""</text>
    </message>
    <message line="7">
      <author>b61c86937f29437dba66cae2be8a9734</author>
      <time>20:32</time>
      <text>13??</text>
    </message>
    <message line="8">
      <author>001744005608bb20b997db6d8cabb3a9</author>
      <time>20:32</time>
      <text>yea</text>
    </message>
```

Figure 3.2: Structure of XML file.

Many people thought it sounded too comprehensive and like too much work for them to want to participate. Further, several of the volunteering participants ended up doing very little or nothing, resulting in less evaluations of conversations than we initially were hoping for. Out of the 36 initially signed up for participation, only 20 did actually participate. The number of evaluations each participant contributed with ranged from a few to dozens.

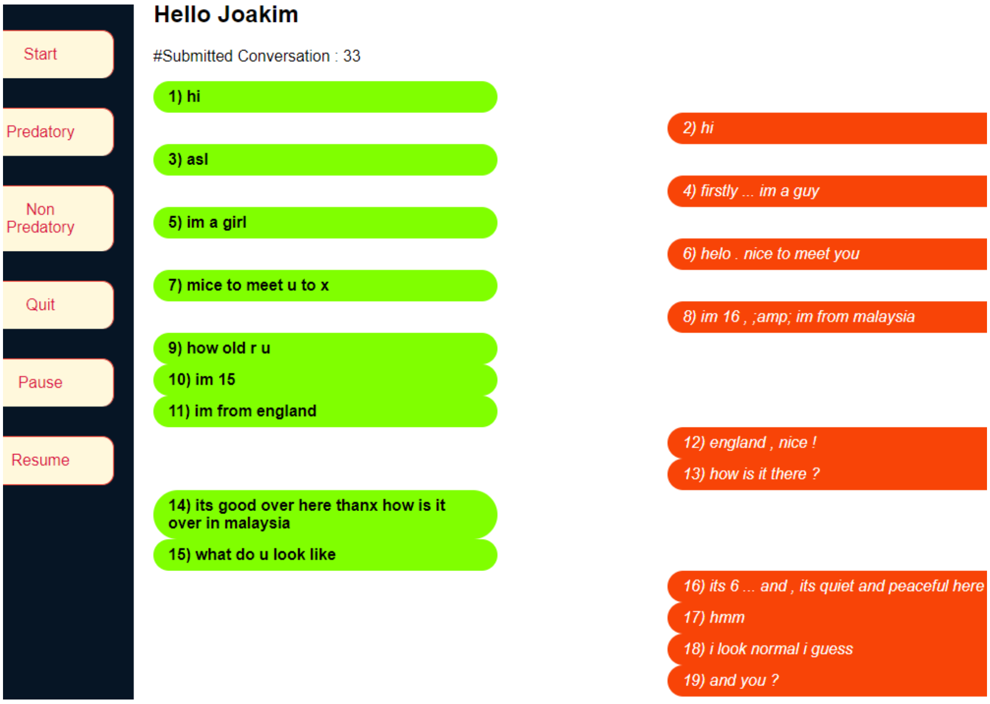
3.2.2 The Experiment

The experiment used for the collection of human evaluation of conversations for this thesis was conducted through a web application online. The web application was provided by the supervisor of this thesis, and created for the specific purpose at NTNU Gjøvik. This allowed the participants to participate in the comfort of their own surroundings and at a time that suited their schedule the best. It also limited potential spread of COVID-19, as it was not necessary to gather people in one location.

For the web application, a user account was created for each user in order to keep track of gender and age. This also allowed each user to log in and out as many times as they wanted, in the hope that they would do more evaluations over a period of time by doing some now and some then.

Figure 3.3 is a screenshot of the web application used for the experiment. On the top, it greets the user and shows how many submitted evaluations the person have in total. On the left hand side, it has a menu with action buttons; "Start", "Predatory", "Non-Predatory", "Quit", "Pause" and "Resume".

The button "Start" starts a new conversation for evaluation. A conversation is equal to one XML file in the dataset. The conversation is then displayed message by message with a few seconds in between, in chronological order. Each party of the conversation is represented by its own color in the main field of the screen, to the right for the menu. The first one to write a message is represented by green on the left hand side, and the other party is represented with red color on the right hand side. When a user has read enough messages to evaluate the conversation to be potentially predatory or non-predatory, the buttons "Predatory" or "Non-Predatory" are used respectively. When the button "Non-Predatory" is clicked, the conversation stops and a dialog box pops up on the bottom right side of the screen as shown in figure 3.4. From this dialog box the user uses radio buttons to select if the conversation is sexual or normal and writes a few words explaining the decision before hitting the "Submit" button. After the submission, the user is presented to the rest of the conversation as shown in figure 3.5. The user can then read through the remaining of the conversation and decide whether to stand by the made decision by hitting the "Continue" button on the left side, or to change the decision by hitting the "Change Decision" button on the left side. When a user wants to change decision, a new dialog box is shown on the top of the screen prompting the user for a reason to why he/she wants to change the decision. When the reason is given, the user is then given the option to choose predatory or



Hello Joakim

#Submitted Conversation : 33

1) hi

2) hi

3) asl

4) firstly ... im a guy

5) im a girl

6) helo . nice to meet you

7) mice to meet u to x

8) im 16 , ;amp; im from malaysia

9) how old r u

10) im 15

11) im from england

12) england , nice !

13) how is it there ?

14) its good over here thanx how is it over in malaysia

15) what do u look like

16) its 6 ... and , its quiet and peaceful here

17) hmm

18) i look normal i guess

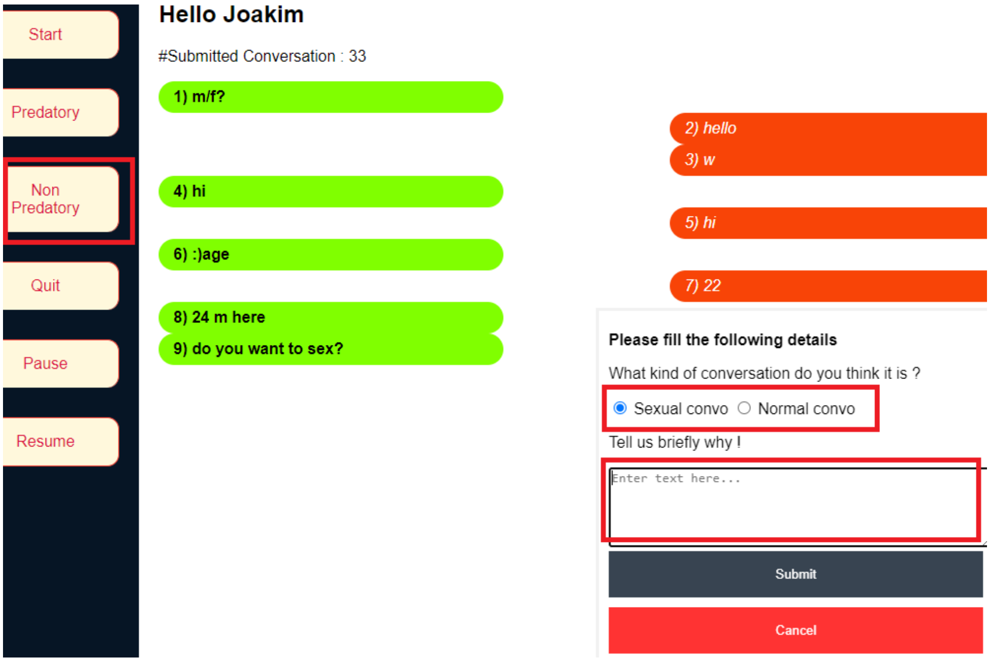
19) and you ?

Figure 3.3: The graphical user interface of the experiment (GUI)

non-predatory again. For cases where the user thinks the conversation is potentially predatory, the "Predatory" button in the menu is used. The conversation is once again stopped, and a dialog box pops up on the bottom right side, similar to the dialog box for non-predatory. What is different with this dialog box, is that the user will have to choose what side he/she thinks is the predator by selecting one of two radio buttons stating "The left one (green)" and "The right one (red)". Below the radio buttons, the user then describes with a few words or sentences why they came to the conclusion. Next the "Submit" button is hit to submit. The remaining of the conversation is then displayed in full and the user can read through it and decide whether or not he/she will stand by the made decision or if it is necessary to change decision, in the same way as with non-predatory conversations. By clicking "Continue" in the menu on the left hand side the evaluation is finished and submitted, and by clicking "Change Decision" a dialog box shows on the top of the browser prompting for a reason to why decision is to be changed. The user will then get back to the menu where predatory or non-predatory can be chosen over again.

3.2.3 Data Result from Experiment

From the experiment, a lot of valuable data was collected from human evaluations. From the database of the web application, a CSV file was exported containing data collected for each evaluation by each user. The exported CSV file consists



Hello Joakim

#Submitted Conversation : 33

1) m/f?

2) hello

3) w

4) hi

5) hi

6) :)age

7) 22

8) 24 m here

9) do you want to sex?

Please fill the following details

What kind of conversation do you think it is ?

● Sexual convo ○ Normal convo

Tell us briefly why !

Enter text here...

Figure 3.4: Experiment GUI: A conversation is marked as non-predatory and sexual. A few words is to be written before the submission.

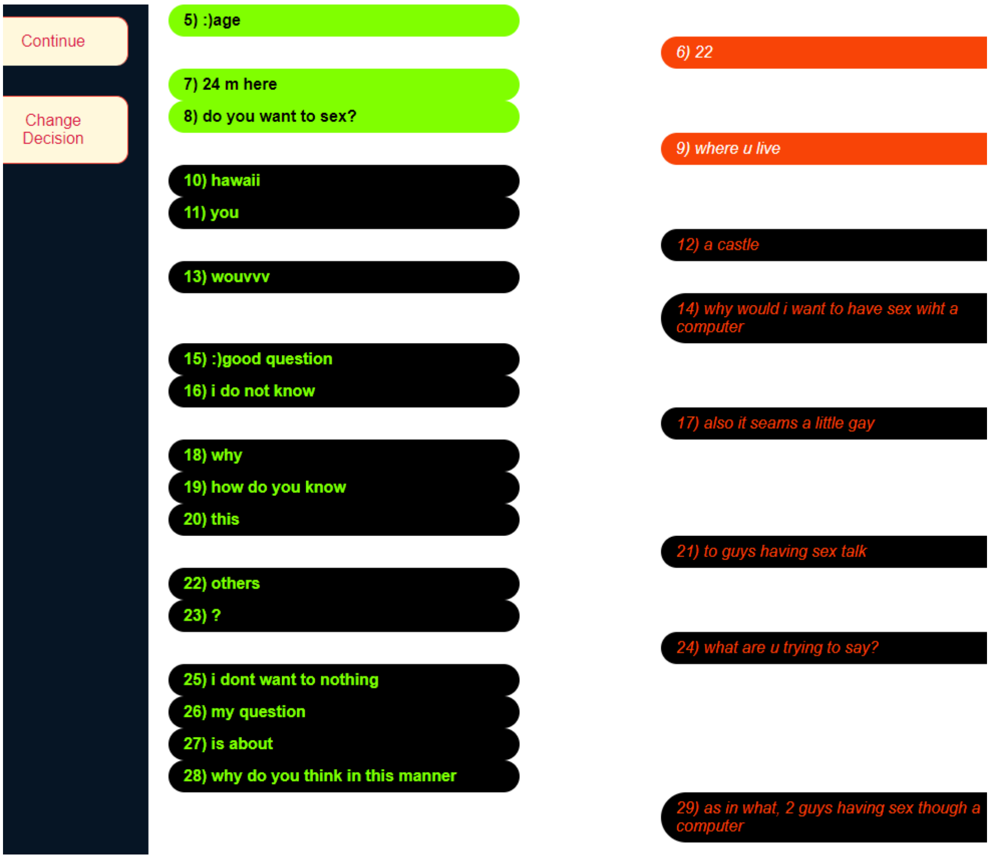
of 10 columns. The different columns consists of the following:

1. Index counter.
2. File ending number, part of XML file name for the specific conversation.
3. How many messages back and forth the participant needed before a decision was made.
4. Participant ID
5. Analysis result; predatory or non-predatory
6. Subresult of analysis result; left or right for predatory conversations and normal or sexual for non-predatory conversations.
7. Date and time for when the decision was made.
8. Final (1) or changed (0) decision. In case of changed decision (0), the following row will give the changed decision.
9. Text field with reason for the participants decision.
10. Dataset name, part of XML file name for the specific conversation.

By combining column 10 and 2, we get the exact file name for the specific XML file for the conversation in question.

3.3 Data Preparation

In order to use the data from the datasets it was necessary to do some preparations and preprocessing, as there were several data sources that needed to be combined



5) :)age

6) 22

7) 24 m here

8) do you want to sex?

9) where u live

10) hawaii

11) you

12) a castle

13) wouvvv

14) why would i want to have sex wiht a computer

15) :)good question

16) i do not know

17) also it seams a little gay

18) why

19) how do you know

20) this

21) to guys having sex talk

22) others

23) ?

24) what are u trying to say?

25) i dont want to nothing

26) my question

27) is about

28) why do you think in this manner

29) as in what, 2 guys having sex though a computer

Figure 3.5: Experiment GUI: The remaining messages of the conversation is displayed to the user after submission.

in order to get meaning from the data.

A python script was made using the libraries lxml and Pandas. The lxml library was used to get powerful tools for parsing the XML files, and the Pandas library was used to store, keep track of and utilize all the data.

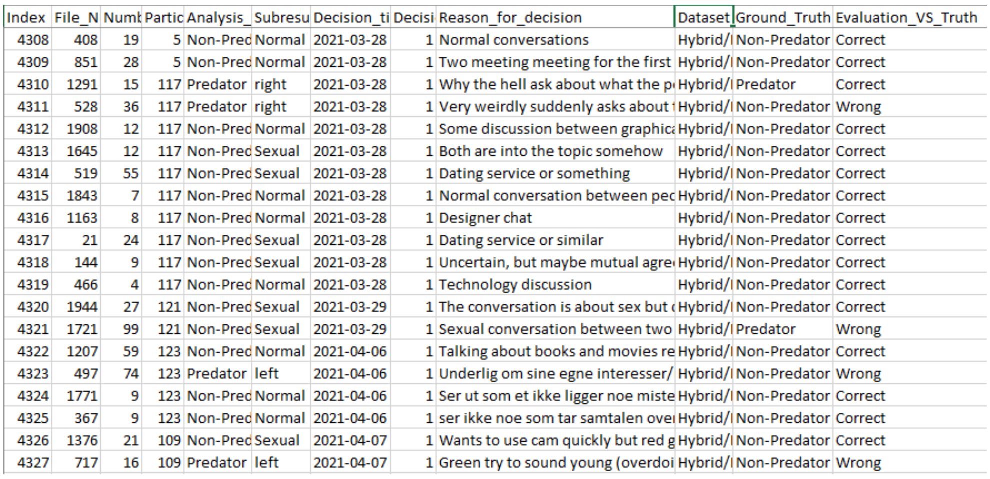
Each XML file of the dataset containing 32063 files from Bours and Kulsrud were first parsed through to find the file number, conversation ID and the author IDs of the conversation. Each file was stored as a new data row in a Pandas DataFrame, where the different values were stored in their own columns. When files were parsed, the first author ID of the conversation, meaning the first person to send a message, was checked against the ground truth file. If the ID was found, the author ID was stored in the DataFrame. The subresult column of the DataFrame were labeled left for the file in question, the conversation was labeled predatory and the script proceeded to the next file. If the first ID of the conversation was not found, the second one was checked against the ground truth file. If this second ID was found, it was labeled right. The conversation was labeled predatory and the script proceeded to the next file. If none of the two author IDs of the conversation were found, the file was labeled non-predatory and the script proceeded to the next file. The labeling of left and right is a reference for visual representation of the conversations, just as the conversations are displayed in the experiment GUI. The left hand side is always the first party to send a message, and the right hand side is always the other party of the conversation. The labeling of predatory (left and right) and non-predatory also makes it easier to process and analyse the data at a later point of time.

After the processing of the 32063 files, the dataset of 4084 files was processed in the same way and the data stored in a new DataFrame. In order to find what files in the 32063 dataset the 4084 manually evaluated files corresponds to, the conversation IDs stored for each conversation in the two DataFrames were compared. For those matching, the filename and number from the 32063 dataset was added to a new column for the file in question from the 4084 dataset.

Next, the processed data was combined with the output data from the experiment in order to see how the different evaluations were compared to the ground truth. Initially, the data from the experiment was without ground truth, only holding the information given from the evaluation. Therefore, to be able to analyze the data, one necessary part of it was to link it up with the ground truth. This was achieved by extending the python script to import the evaluations data into another DataFrame. This was used to match the file name and numbers of the evaluation data with the data stored in the DataFrame for the 4084 dataset holding information about ground truth. In addition to adding a column with the ground truth, another column was added comparing the evaluation to the ground truth to ease the process of finding evaluations deviating from the ground truth.

Finally, the new DataFrame holding the evaluations data combined with the ground truth and comparison was exported into a CSV file, which figure 3.6 shows an excerpt of.

The CSV file was now ready for analysis and all preprocessing was completed.



Index	File_N	Numb	Partic	Analysis_	Subresu	Decision_ti	Decisi	Reason_for_decision	Dataset_	Ground_Truth	Evaluation_VS_Truth
4308	408	19	5	Non-Pred	Normal	2021-03-28	1	Normal conversations	Hybrid/I	Non-Predator	Correct
4309	851	28	5	Non-Pred	Normal	2021-03-28	1	Two meeting meeting for the first	Hybrid/I	Non-Predator	Correct
4310	1291	15	117	Predator	right	2021-03-28	1	Why the hell ask about what the p	Hybrid/I	Predator	Correct
4311	528	36	117	Predator	right	2021-03-28	1	Very weirdly suddenly asks about t	Hybrid/I	Non-Predator	Wrong
4312	1908	12	117	Non-Pred	Normal	2021-03-28	1	Some discussion between graphica	Hybrid/I	Non-Predator	Correct
4313	1645	12	117	Non-Pred	Sexual	2021-03-28	1	Both are into the topic somehow	Hybrid/I	Non-Predator	Correct
4314	519	55	117	Non-Pred	Sexual	2021-03-28	1	Dating service or something	Hybrid/I	Non-Predator	Correct
4315	1843	7	117	Non-Pred	Normal	2021-03-28	1	Normal conversation between pec	Hybrid/I	Non-Predator	Correct
4316	1163	8	117	Non-Pred	Normal	2021-03-28	1	Designer chat	Hybrid/I	Non-Predator	Correct
4317	21	24	117	Non-Pred	Sexual	2021-03-28	1	Dating service or similar	Hybrid/I	Non-Predator	Correct
4318	144	9	117	Non-Pred	Sexual	2021-03-28	1	Uncertain, but maybe mutual agre	Hybrid/I	Non-Predator	Correct
4319	466	4	117	Non-Pred	Normal	2021-03-28	1	Technology discussion	Hybrid/I	Non-Predator	Correct
4320	1944	27	121	Non-Pred	Sexual	2021-03-29	1	The conversation is about sex but (	Hybrid/I	Non-Predator	Correct
4321	1721	99	121	Non-Pred	Sexual	2021-03-29	1	Sexual conversation between two	Hybrid/I	Predator	Wrong
4322	1207	59	123	Non-Pred	Normal	2021-04-06	1	Talking about books and movies re	Hybrid/I	Non-Predator	Correct
4323	497	74	123	Predator	left	2021-04-06	1	Underlig om sine egne interesser/	Hybrid/I	Non-Predator	Wrong
4324	1771	9	123	Non-Pred	Normal	2021-04-06	1	Ser ut som et ikke ligger noe miste	Hybrid/I	Non-Predator	Correct
4325	367	9	123	Non-Pred	Normal	2021-04-06	1	ser ikke noe som tar samtalen ove	Hybrid/I	Non-Predator	Correct
4326	1376	21	109	Non-Pred	Sexual	2021-04-07	1	Wants to use cam quickly but red g	Hybrid/I	Non-Predator	Correct
4327	717	16	109	Predator	left	2021-04-07	1	Green try to sound young (overdoi	Hybrid/I	Non-Predator	Wrong

Figure 3.6: Excerpt from the CSV file generated by the python script.

The analysis will be described in the next chapter.

Chapter 4

Analysis and Results

This chapter explains the analysis performed on the data collected from the experiment and the summer intern and the results from it.

The analysis process on the evaluations and the corresponding conversations was performed to get a better understanding of what features makes conversations predatory or non-predatory. A machine learning model has limited abilities to understand conversations, as it only recognizes certain types of repeating patterns. By analysing evaluations and conversations manually, it is possible to discover patterns a machine learning model is not able to find. It is also possible to find single features humans find to be describing and useful in detection of potentially predatory conversations.

When talking about the different conversations and evaluations from the different datasets we will from now refer to the evaluations and conversations performed by the summer intern as PAN and the ones gathered from the data collection experiment as Hybrid.

The prepared CSV file containing data from the experiment and the ground truths, plus the conversations themselves were the basis for the analysis. As a comment was made for each of the evaluations, the analysis process aimed to derive meaningful trends and patterns from the quantity of multiple evaluations of many conversations. As the existing system today is created to only evaluate the risk of each message isolated, it lacks the ability to address more complex features of conversations that can be used to detect potential predatory conversations. The analysis aims to discover such features.

4.1 Analysis Method

The analysis was performed as a qualitative study, utilizing the powers of content analysis. Content analysis was used to gain insight and discover trends and patterns from evaluations and conversations in order to extract value from human knowledge to be utilized towards the making of a hybrid cyber grooming detection system. Content analysis is known to be a good research approach in cases where human communication and interaction is to be investigated. This made it highly valuable for this purpose as the raw data exclusively consists of chat conversations between two people and the collected evaluation data is based on the conversations [33].

Content analysis was used in two different ways for the purpose of this thesis, which combined resulted in the analysis and results. First, the evaluations gathered from the data collection experiment was analyzed to discover what the human evaluators did react to when evaluating the conversations. Second, for each evaluation the corresponding conversation was analyzed in order to better understand the evaluations and to find more describing features of conversations. The findings from the combinations of these two are presented in text and bullet points under their respective datasets and evaluated classes.

In order to gain extensive knowledge and insight from the evaluations and conversations, a large number of the evaluated conversations were manually analyzed for the writing of this thesis. Manual analysis in such context can have subjective bias, which is important to consider, and which was present in this case. Also the different evaluations from the data collection contains bias in different directions based on the different evaluators experience, knowledge and thoughts. This bias was, however, leveled as good as possible by having several evaluators and not only one or a few. In regards of the bias from the manual review, the reviewer possesses in-depth knowledge on the topic of cyber grooming and cyber grooming detection at a much higher level than most people. This can lead to subjective findings which are biased towards already known knowledge on the topics, and potentially limit openness to discover new findings never previous highlighted. Even though subjective findings in general are something not as wanted as objective findings, it can be argued that subjective findings in regards of cyber grooming detection can be as valuable as, or close to as valuable as, objective findings. Because there is no correct answer to what defines a potentially predatory conversation, subjective findings can either serve as stand-alone features or contribute with additional value to objective findings towards a hybrid cyber grooming detection system [33].

In order to equalize the bias as good as possible, the results from the manual reviews were seen in the context of the evaluations of the conversations from the data collection. Not only was this done to reduce potential bias, but also to add additional value to the data basis for this thesis [33].

4.2 Evaluations and Conversations

From the datasets of conversations, there were in total 4578 unique evaluations of conversations, of which some of them were of the same conversations. Out of those, 4084 were evaluated from the PAN dataset (in the CSV file shown in figure 3.6 referred to as dataset "PAN12/Conversation_"), and 494 were evaluated from the Hybrid dataset of 2000 conversations (in the CSV file shown in figure 3.6 referred to as dataset "Hybrid/HybridConv_"). The 4084 were evaluated by the summer intern alone. On the reduced Hybrid dataset of 2000 conversations, the 494 evaluations were performed by participants in the experiment. Table 4.1 shows a visual representation of how many evaluations was performed on each dataset.

Dataset	**Number of evaluations**
PAN	4084
Hybrid	494
Total	4578

Table 4.1: Number of evaluations on the different datasets.

Table 4.2 shows that out of the evaluated conversations, 280 (6,86%) of the conversations from the PAN dataset are defined actually predatory by the ground truth and 23 (4,66%) of conversations from the Hybrid dataset. The remaining 3804 (93,14%) from the PAN dataset and 471 (95,34%) from the Hybrid dataset are defined as non-predatory by the known ground truth. This shows that the proportions of predatory conversations are percentage wise just above what it is in the original PAN-2012 dataset [31], making the dataset characteristics approximately the same for the datasets used for this thesis as for the original PAN-2012 dataset. It could, however, be argued that the proportion of predatory conversations should have been larger for the purpose of this thesis data collection in order to get more evaluations of actually predatory conversations. But, as the analysis shows, there is a lot of value also found in evaluations of conversations defined non-predatory by default. Because the existing detection system uses classification, it is already able to find features on its own from conversations classified as non-predatory. But, it is not able to find potentially predatory features from conversations defined as non-predatory. This because the system treats all these conversations as non-predatory and harmless due to how machine learning classification works. In other words, the number of actually predatory conversations present in the datasets will make the number of non-predatory conversations evaluated larger. This potentially calls for more false positive evaluations. Such false positives have the advantage of not being staged, i.e. the potentially predatory conversations discovered and evaluated are probably of predators communicating with real victims, and not volunteers posing as victims. This allows for discovery of features present in more realistic conversations, as one of the sides not are trying to provoke predatory actions from the opposite side.

Dataset	**Predatory**	**Non-predatory**
PAN	280	3804
Hybrid	23	471

Table 4.2: Characteristics of the evaluated conversations based on the known ground truth for the different datasets.

4.3 Hybrid

From the 494 evaluations gathered on the Hybrid dataset, 23 (4,66%) conversations are actually predatory and 471 (95,34%) non-predatory. From the evaluations, 9 (1,82%) were evaluated True Positive (TP), 61 (12,35%) False Positive (FP), 410 (83,00%) True Negative (TN) and 14 (2,83%) False Negative (FN). These characteristics from the evaluations are presented as a confusion matrix in table 4.3.

		Predicted Class	
		Positive	Negative
True Class	Positive	9 (1,82%)	14 (2,83%)
	Negative	61 (12,35%)	410 (83,00%)

Table 4.3: Confusion matrix representing characteristics of the evaluations on the Hybrid dataset

4.3.1 True Positives Hybrid

The true positives evaluations, i.e. those correctly evaluated to be predatory, provides valuable insight on how to define predatory conversations and what features about real predatory conversations humans do react to. Because these conversations are predatory, the evaluations of these conversations are therefore extremely valuable as each evaluation and conversation feature pointed at will be part of a harmful conversation.

From the Hybrid dataset there are a small number of true positives, only 9, which ideally should have been larger. Due to this fact, trends and patterns can not be established in the same way as if there were a bigger number of true positives. However, since each of the true positive evaluations are of actual predatory conversations, the value they represent is greater than the value from false positive evaluations. Because the true positives contains actually predatory conversations, single evaluations and analysis conversations can be weighted heavier than for example with false positives.

By manually reviewing the conversations, it is possible to further support features highlighted by evaluations. Several of the predatory conversations are of a sexual or semi-sexual art, where the predatory side is the leading part asking most of the questions and driving the conversation forward, as figure 4.1 is an example of. Some cases are straight to the point where the predator talks about what sexual things he/she would like to do to the victim or how the victim would like it if the predator is to perform sexual actions to him/her. In other cases the predator are more careful where the predator e.g. starts asking what the victim is wearing, if the victim would like the predator to be there and what the victim would like to do to the predator when they meet. Such cases also develop further into more ongoing behavior from the predator, sometimes ending up with more direct sexual behavior where the predator e.g. says he/she wants to perform aggravated sexual actions to the victim. Also, the predatory side tends to have a more well developed and sometimes rougher language, in addition to being more ongoing than the victim.

Predators are also open about their age, as shown in the example of figure 4.1. The very same example also poses as an example of a predator making sure the victim is alone before initiating a meeting.

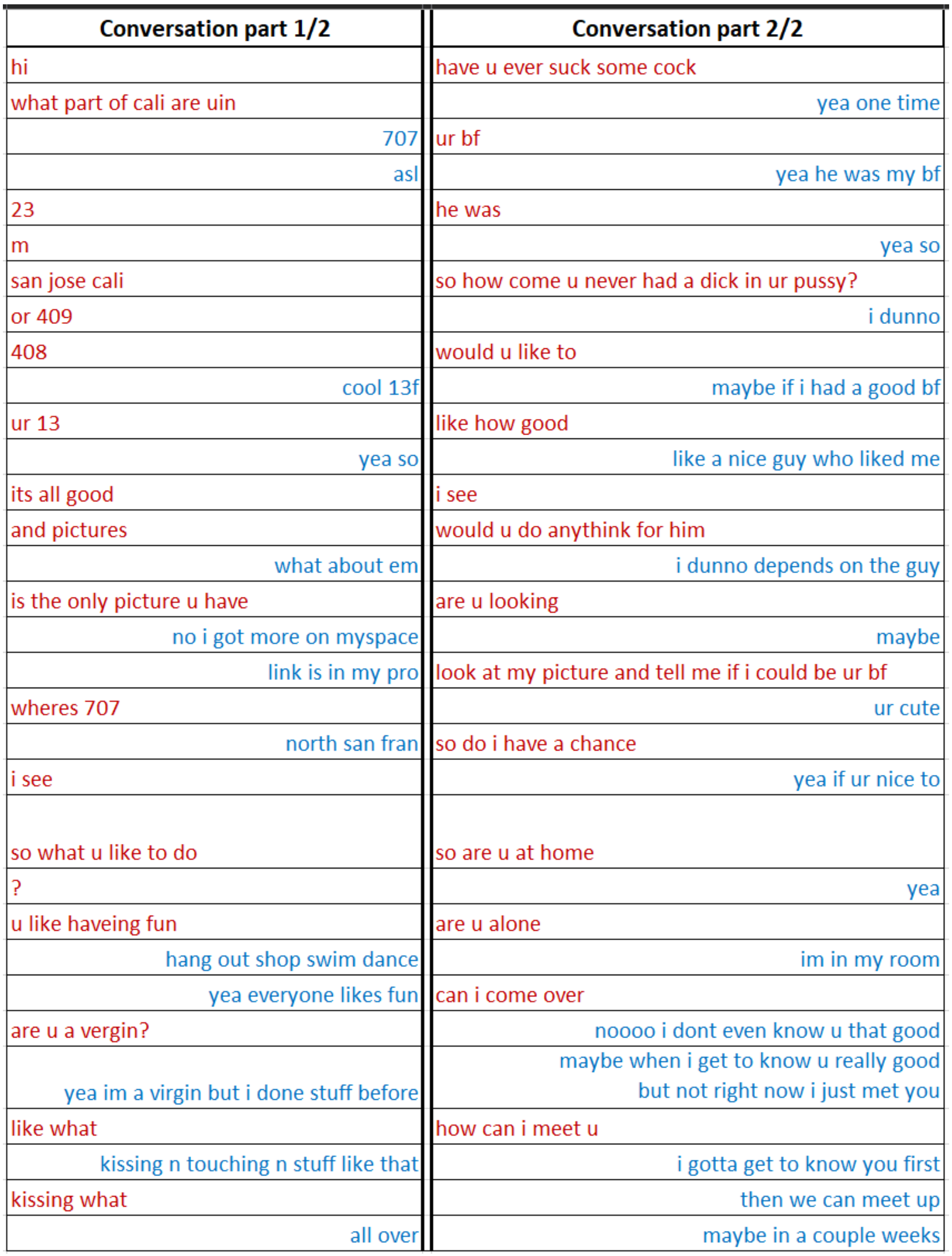
Features from conversations correctly identified as predatory:

- Conversations are sexual.
    - Sexual wants and needs.
    - If the victims would like the predators to perform different sexual actions to them.
    - What sexual actions the victims would like to do or fantasizes about.
- Victims are afraid sexual actions will hurt.
- One of the chatters appears to have a more developed written language than the other party.
- Predators asks what victims are wearing.
- Predators are eager to find out if victims are alone.
- Predators are sometimes more ongoing than victims.
- Predators are often quite eager to arrange and schedule meetings.
- Predators are often open about their age.
- Predators asks for pictures.
- Predators asks more questions than victims.

4.3.2 False Positives Hybrid

False positive evaluations are those evaluated to be predatory, which actually are classified as non-predatory by the ground truth.

The majority of the false positive evaluations are evaluated wrong (i.e. to be predatory when actually non-predatory due to the ground truth) because of the dataset they originates from. By default, all of these conversations are classified as non-predatory because of where they are collected from. Even though most of them are actually non-predatory, it is possible that some of the conversations actually are potentially predatory. But, this is something we do not know, and we do not know if any of them have been subject for any legal prosecution. What is possible to determine however is that some of the conversations are between one conversation party stated to be underage and the other one stated to be above legal age, making them potentially predatory in some cases. Such cases are for instance when the potentially predatory party of the conversation either initiates a meeting or agrees to a meeting with the intention of performing sexual abuse to the victim. Since it is impossible to determine what intentions someone has, the fact that an adults plans a meeting with an underage child would be enough to at least flag a conversation in order for it to get attention from e.g. a moderator. This



Conversation part 1/2	Conversation part 2/2
hi	have u ever suck some cock
what part of cali are uin	yea one time
707	ur bf
asl	yea he was my bf
23	he was
m	yea so
san jose cali	so how come u never had a dick in ur pussy?
or 409	i dunno
408	would u like to
cool 13f	maybe if i had a good bf
ur 13	like how good
yea so	like a nice guy who liked me
its all good	i see
and pictures	would u do anythink for him
what about em	i dunno depends on the guy
is the only picture u have	are u looking
no i got more on myspace	maybe
link is in my pro	look at my picture and tell me if i could be ur bf
wheres 707	ur cute
north san fran	so do i have a chance
i see	yea if ur nice to
so what u like to do	so are u at home
?	yea
u like haveing fun	are u alone
hang out shop swim dance	im in my room
yea everyone likes fun	can i come over
are u a vergin?	noooo i dont even know u that good
yea im a virgin but i done stuff before	maybe when i get to know u really good but not right now i just met you
like what	how can i meet u
kissing n touching n stuff like that	i gotta get to know you first
kissing what	then we can meet up
all over	maybe in a couple weeks

Figure 4.1: Example of a predator open about age, turning the conversation sexual and initiates meeting.

is one important reason to why false positive evaluations are of great value even though they are not classified predatory in the PAN-2012 dataset. It is, however, important to note that it is also possible for underage people to impersonate older people. There could be multiple reasons for why someone would do this, for example if the impersonator thinks it is more attractive if he/she appears to be an older person.

A large number of these evaluations are evaluated to be predatory due to the fact that they are sexual without any further stated considerations. A sexual conversation in itself is not enough for a conversation to be predatory. Several of the conversations showed from the manual review to be between chatters above legal age (however, some were 18 years old or younger), whereas several of them also were sexual. This indicates that the evaluations have been performed and based on either the though that victims are persons under the age of 18 years old instead of under the age of 16 years old, or the thought that a conversation is predatory if it is sexual. These assumptions are further supported by the evaluations as most of them only points at certain characteristics of the conversations, e.g. sexual, impersonation, well developed language and other characteristics. As conversations are non-predatory if the victim is 16 years or older, conversations where both chatters are above 16 years old are not illegal in any way in Norway and the predatory side cannot be prosecuted unless any other criminal laws have been violated. However, false positive evaluations from conversations where the assumed victim side is just above legal age or up to around 18 years old can hold certain value. Even if the assumed victim side is above legal age, the predatory sides pattern of action holds a certain value, especially if there are some age gap between the predatory side and the victim side. This because the action patterns of such conversations potentially can be very much like in other conversations where the victims actually are underage. Therefore, such conversations and evaluations of them are great to include when looking at behavior, trends and patterns of potentially predatory conversations.

Some of the conversations are also evaluated false positive without any reasonable substance. By reasonable substance in this case we mean incomplete comments, vague comments, or comments pointing at features which are to be considered not defining a potentially predatory conversation. For example one side of a conversation is swearing. This is a feature that does not hold as evidence of a potentially predatory conversation because it is something that can be done by anyone and, even though it might be inappropriate, it is a fairly normal feature of peoples language. All of the conversations were also manually reviewed, which allowed to confirm some of the conversations to be without reasonable substance, as they were without any reasonable evidence of being potentially predatory. In other words, there were no indicators of any of the sides being underage, no sexual writing, no talk about meetings, and generally just ordinary conversations with sometimes odd features.

Features from conversations evaluated false positive:

- One of the chatters is underage and the other party above legal age.

- Conversations are sexual with one of the chatters being underage.
- Impersonation.
    - One of the parties have a more developed language than what to expect from age stated by the impersonator.
- Predators sometimes have a better developed written language.
- Predators sometimes adapt their written language to match the victims language.
- Predators are eager to establish multiple points of contact by trying to get MSN, phone number or other ways of communicating with the victims.
- One party of the conversation quickly wants to get pictures of the other conversation party, in some cases also nude pictures.

4.3.3 True Negatives Hybrid

True negative evaluations are those evaluated to be non-predatory, which also are classified as non-predatory by the ground truth. These evaluations counts for the majority of the evaluations gathered from the data collection, which also correctly reflects the selection of different types of conversations in the datasets.

The vast majority of the true negative evaluated conversations are actually non-predatory conversations which appears normal. The definition of normal is hard to describe, as what is considered as normal is a subjective manner depending on the person defining it. It is however possible to shape a rough outline from the evaluations. A normal conversation in this context most often consists of information sharing where two people, which often are strangers to each other, are trying to ask questions about each other and answering questions. This in a way they can get to know what kind of persons they are. Typical topics and information shared contains information about age, gender, place of living and hobbies. Figure 4.2 shows an example of a normal conversation where the two conversation parties are sharing basic information in order to get to know each other. Another slight different normal is when the two conversation parties apparently know each other from before and the relationship already is established. In these situations, normal often consists of questions related to the well being of the opposite party, e.g. by asking what the other person has been doing today or if he/she is doing well. Also, in these situations hobbies or interests are frequently discussed as part of normal, in addition to entertaining conversations about different topics.

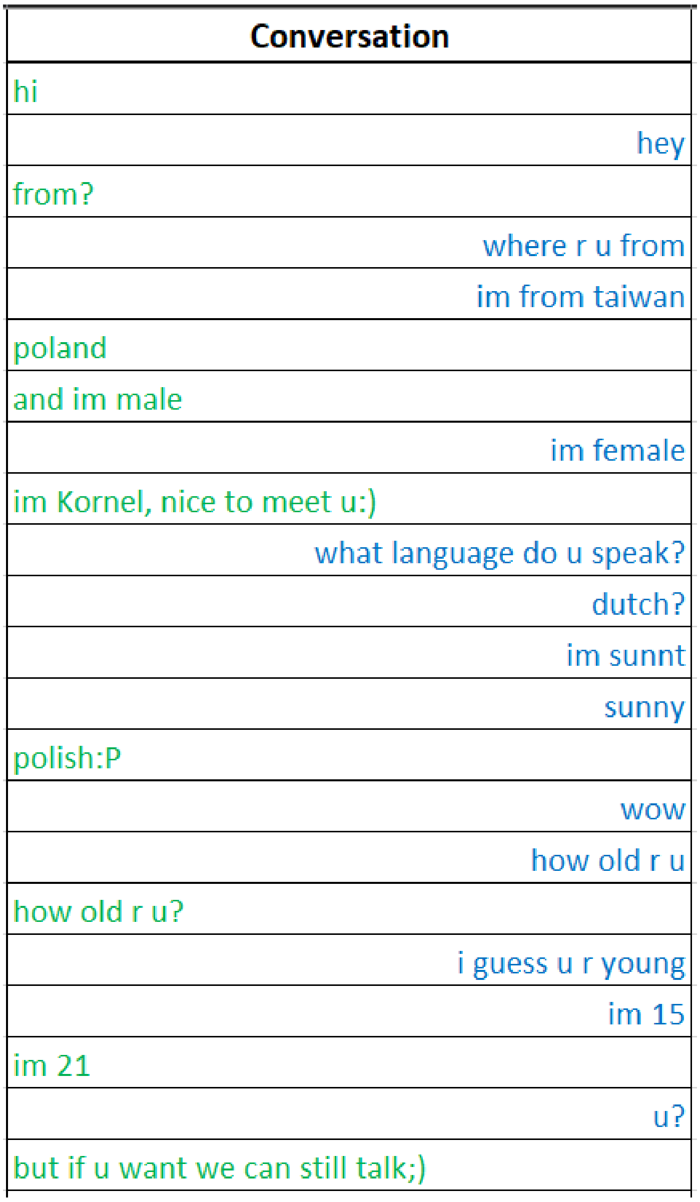
When it comes to age, the conversations are mostly either between two conversation parties both above legal age, or both below. Age is not always written out in plain text, but in many cases it is, as shown in figure 4.2.

Only 8 conversations evaluated true negative at first was changed into false positive after the participants got to read through the whole conversation, which corresponds to 1,95% of the true negatives. This a very low number and can indicate that if a conversation starts off as supposedly non-predatory, it is highly likely that it also is non-predatory. Features that help establish this assumption from conversations are introduction of age early in conversations, mutual interest in each other and approximately equal level of written language level. It is still important to note that there are conversations turning out to be predatory even after seemingly harmless behavior, which makes it impossible to use this alone to write off conversations from being potentially predatory.

There are several sexual conversations evaluated true negative from the data collection. What is common for all of these is that they are not harmful as they are between adults or people above legal age. Sexual conversations or conversations with a sexual undertone shows to appear quite frequently as 20% of the true negatives are sexual, which corresponds to every fifth conversation. We know from evaluations of actual predatory conversations that they tend to be sexual (at some point), and it can be one way of helping identifying a potentially predatory conversation. The fact that also non-predatory conversations also quite often appear to be sexual then complicates detection of potentially predatory conversations based on sexual vs. non-sexual content.

Features from conversations evaluated true negative:

- Normal conversations.
- Conversations about technical topics often containing very technically specific terms.
- Adults or people above legal age chatting.
- Some conversations are sexual.
    - Cyber sex.
    - Sexual preferences.
    - Sexual wants and needs.
- Most conversations are non-sexual.
- Written language is often of relatively equal level.
- Some conversations are ordinary talks between people trying to get to know each other.
    - Between both young people and adult people.
    - Sharing basic information.
    - Sharing contact details.
    - Discussing hobbies and common interests.
- Conversations are entertaining with funny content.
- Ordinary smalltalk on all sorts of topics.
    - Sports.
    - School.
    - Work.
    - Hobbies.
    - Interests.



Conversation
hi
hey
from?
where r u from
im from taiwan
poland
and im male
im female
im Kornel, nice to meet u:)
what language do u speak?
dutch?
im sunnt
sunny
polish:P
wow
how old r u
how old r u?
i guess u r young
im 15
im 21
u?
but if u want we can still talk;)

Figure 4.2: Example of a non-predatory normal conversation between two young people getting to know each other by asking questions and sharing information.

4.3.4 False Negatives Hybrid

False negative evaluations are those evaluated to be non-predatory, which actually are classified as predatory by the ground truth.

The analysis of the evaluations and further the conversations were highly valuable as it provided some additional insight to better understand why supposedly predatory conversations were evaluated as non-predatory. Two key elements stood out to shape a pattern; most conversations were non-sexual and only 3 of the conversations were sexual, making it 21,43% of the conversations. The other key element is that chatters obviously knew each other from before these conversations. This means the conversations must have been continuations of previous conversations. Figure 4.3 shows an example where the chatters obviously knows each other from before. We do know from the creation process of the PAN-2012 dataset that conversations have been split up when there was a pause in communication for 25 minutes or more, which makes the assumption that the chatters knew each other highly probable.

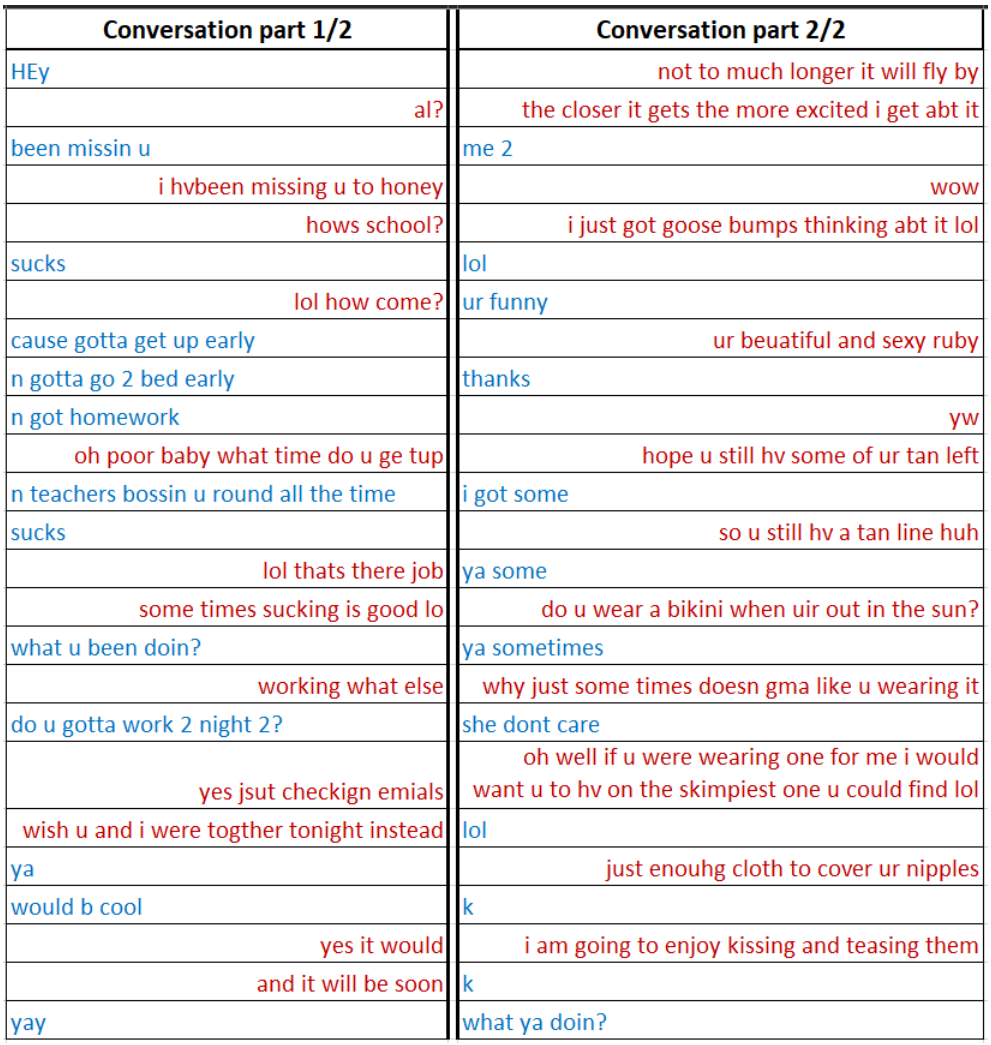
Even though it is hard or impossible to determine the chatters age, it is to some extent possible to assume one chatter to be underage and the opposite chatter to be above legal age based on different aspects of the conversations. Such aspects are e.g. one side talking about parents at home, and the other chatter talking about drinking alcohol, driving a car or going to work. The example in figure 4.3 contains enough information about both chatters to assume the left side of the conversation to be a child in school, and the right side to be an adult due to the fact that he/she is working at night.

None of the decisions on the evaluations were changed after the participants got to read through the whole conversation. This can potentially indicate one of two things. Either the participants were too hung up on the thought that the conversations needed to be of a sexual character for them to be considered predatory, an assumption that can be supported by some of the evaluation comments stating the conversations were not sexual. Or that the vague signs of age were not discoverable to people without bias and extensive knowledge about predatory conversations and behavior.

Another take away from the false negative conversations is that predators asks a lot of questions, and often it appears to be more than the victims asks. Also, many of the questions asked by predators are related to if the victim is home alone, if any parents are home, where they are, and other similar questions.

Features from conversations evaluated false negative:

- Most conversations have no sexual content.
- Chatters seems to know each other from before the conversation.
- Conversations consists mostly of smalltalk with no clear indicators of predatory actions.
- It is hard or impossible to determine the age of the conversation parties.



Conversation part 1/2	Conversation part 2/2
HEy	not to much longer it will fly by
al?	the closer it gets the more excited i get abt it
been missin u	me 2
i hvbeen missing u to honey	wow
hows school?	i just got goose bumps thinking abt it lol
sucks	lol
lol how come?	ur funny
cause gotta get up early	ur beuatiful and sexy ruby
n gotta go 2 bed early	thanks
n got homework	yw
oh poor baby what time do u ge tup	hope u still hv some of ur tan left
n teachers bossin u round all the time	i got some
sucks	so u still hv a tan line huh
lol thats there job	ya some
some times sucking is good lo	do u wear a bikini when uir out in the sun?
what u been doin?	ya sometimes
working what else	why just some times doesn gma like u wearing it
do u gotta work 2 night 2?	she dont care
yes jsut checkign emials	oh well if u were wearing one for me i would want u to hv on the skimpiest one u could find lol
wish u and i were togther tonight instead	lol
ya	just enouhg cloth to cover ur nipples
would b cool	k
yes it would	i am going to enjoy kissing and teasing them
and it will be soon	k
yay	what ya doin?

Figure 4.3: Example of a sexual, predatory conversation evaluated false negative where the two conversation parties seems to know each other from before.

4.4 PAN

From the 4084 evaluated on the PAN dataset, there were 223 conversations evaluated as True Positive (TP) corresponding to 5,46% of the evaluations. 288 of the conversations were evaluated as False Positive (FP), making up for 7,05% of the evaluations. 57 conversations, 1,40%, were evaluated as False Negative (FN) and 3516, 86,09%, were evaluated as True Negatives (TN).

As all the evaluations on the PAN dataset was performed by one single person, the evaluations are probably not as nuanced as they would have been if there were more people to perform the evaluations. This calls for more subjective findings and findings which potentially can be biased one way or the other. It is also possible that the findings of the evaluator have been shaped over time, i.e. in the beginning some features could have been of great interest and catching attention, whereas after some time and evaluation of multiple conversations the findings could potentially constitute of more features not thought of in the beginning.

		Predicted Class	
		Positive	Negative
True Class	Positive	223 (5,46%)	57 (1,40%)
	Negative	288 (7,05%)	3516 (86,09%)

Table 4.4: Confusion matrix representing the evaluations on the PAN dataset

4.4.1 True Positives PAN

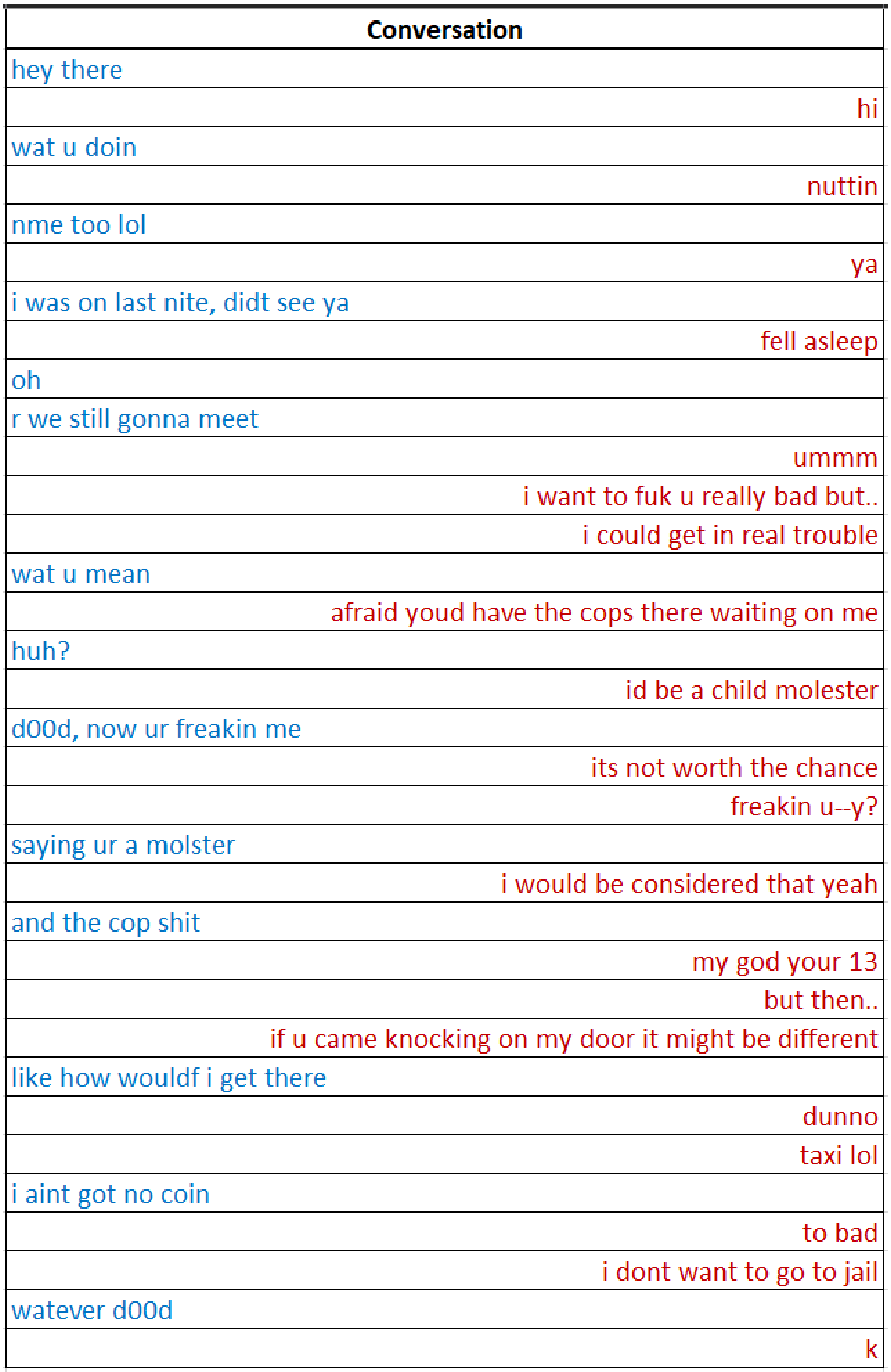
True positives from the PAN dataset evaluations are the conversations evaluated to be predatory which also are classified predatory by the ground truth.

As the number of true positive evaluations is much higher for the PAN dataset evaluations (223 evaluations) than the Hybrid dataset evaluations (9 evaluations), it is easier to discover repeating features, trends and patterns. It is, however, important to notice that the bigger number evaluated from the PAN dataset comes from one single person, which can make the findings subjective and biased towards the evaluators bias.

One significant, repeating feature of conversations evaluated as true positive is the highlighted feature that one of the chatters of the conversations lives at home together with parents, which is the victim part. The age of the victim is also quite clear in most cases from the exchanged information, which emerges from the evaluations where exact age very often is noted in the evaluation comments. As a follow up to this, the manual analysis shows that predators very often asks if those living with the victims are home or if the victims are alone, a statement which also is given on some of the evaluations by the evaluator. This is interesting to see, as it shows that predators are putting a lot of effort into making sure of secrecy. This indicates that predators are perfectly aware of the fact that they are doing something illegal, but it does not stop them, it only takes some more effort for them to hide it as good as possible. In some cases they even state the fact what they are doing is illegal to the victims in order to try get them to tell that they want the predators and that they will stay quiet about their communication. Figure 4.4 shows an example of a predatory conversation where the predator knows what he is doing is wrong. The conversation in figure 4.4 is one out of several conversations which all combined in the end resulted in conviction.

The age of predators are often stated, and predators tend to be surprisingly open about their age. The stated age by predators are making conversations potentially predatory based on the victims stated age. It is, however, possible that some of the predators says they are either younger or older than they actually are. But, this is something which can be quite difficult to detect. In order to potentially be able to find out, other conversation features will have to be considered, and the impersonated story given by the predator must slip at some point.

A lot of the true positive evaluated conversations are sexual, which both figures 4.4 and 4.5 are examples of. This is a repeating trend from the evaluations. The manual review found that the predators most often are the ones to turn the conversations sexual, as seen an example of in figure 4.5. There is not one consistent way predators use to turn conversations sexual, and it varies from conversation to conversation. Some conversations stays just slightly sexual, where the predator e.g. says he/she would like to kiss the victim, cuddle and make love. Figure 4.3 shows an example of a slightly sexual conversation. In other cases they are coarser and more detailed. The way sexual topics are introduced varies from just starting to ask about what clothes the victims are wearing, all the way to straight forward telling the victims they are horny and would like to perform specific sexual actions to them when they meet. What is relatively common is the fact that most sexual conversations sooner or later turns into quite coarse sexual conversations where the content of different conversations becomes a lot more similar. In the beginning, conversations can be quite innocent, but when they turn coarse they can become quite extreme, especially when considering that the victims exposed to this are below 16 years old. Figure 4.8 is an example of a quite coarse conversation, which is an excerpt of a conversation in which some of the messages was removed due to communication issues between the chatters and no real content related to the conversation. Predators often asks if the underage victims would like them to perform all sorts of sexual actions to them, and how they feel about the predators doing it. A lot of what is proposed or said by the predators are sexual actions underage children do know nothing about. In many cases the victims have not even heard of the things proposed, and have in any case not been exposed to. In some cases the predators exploits the emotions of the victims in order to force the victims to say they would like to participate in the actions proposed. Ways of forcing includes the predators telling the victims they are not the person the predators thought or would like if they do not want to participate in the proposed sexual actions. When predators tries to exploit the emotions of victims in order to get them to do or say whatever the predator wants them to, most victims obey. As these sexual actions are unknown territory for the



Conversation
hey there
hi
wat u doin
nuttin
nme too lol
ya
i was on last nite, didt see ya
fell asleep
oh
r we still gonna meet
ummm
i want to fuk u really bad but..
i could get in real trouble
wat u mean
afraid youd have the cops there waiting on me
huh?
id be a child molester
d00d, now ur freakin me
its not worth the chance
freakin u--y?
saying ur a molster
i would be considered that yeah
and the cop shit
my god your 13
but then..
if u came knocking on my door it might be different
like how wouldf i get there
dunno
taxi lol
i aint got no coin
to bad
i dont want to go to jail
watever d00d
k

Figure 4.4: Example of a predator knowing what he is doing is wrong.

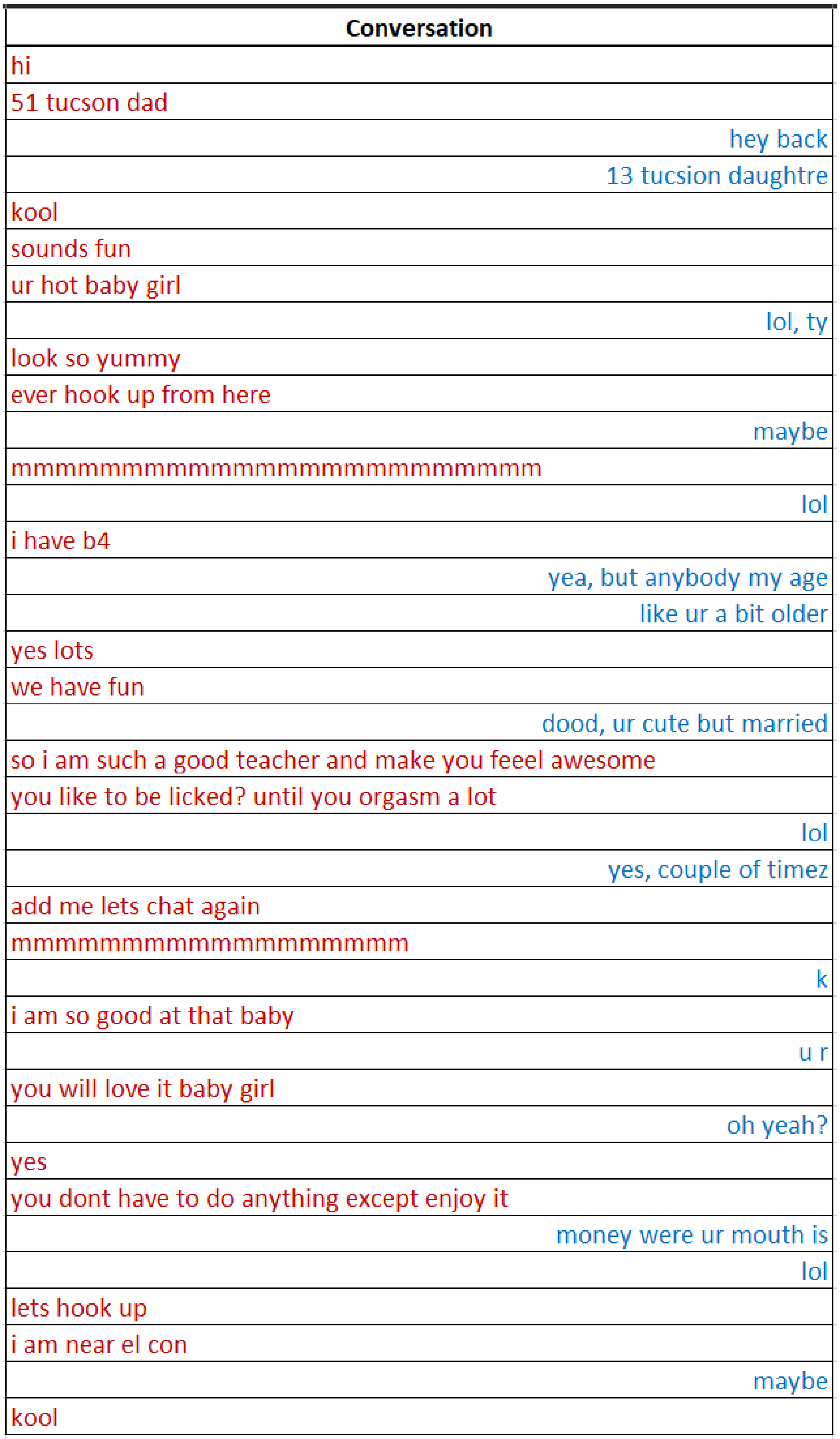
victims, they do in many cases ask if it will hurt and how it works. Such questions supports the statement that these underage victims have previously no or little sexual experience and what the predators wants to do are things most the victims have no knowledge of.

Victims often live at home with their parents or similar. This is sometimes said straight forward upon request from the predators. In other cases, often where the two chatters seem to know each other from previous conversations, the living situation of the victims can sometimes be detected if they are talking about something and states they for example are not allowed by parents, have to wait until their parents are gone, have to do housework before parents gets home, or similar. Figure 4.6 shows an example of a victim revealing the living situation to be with at least his/her mom, as he/she has to do the dishes before mom gets home.

Predators also use a lot of nice words or nicknames for their victims. As figure 4.3 shows, the predator on the right side refers to the victim as "honey".

Features from conversations correctly identified as predatory:

- Predators are often quite eager to arrange and schedule meetings.
- Predators often say "I love you" to their victims.
- Predatory side uses a lot of nice words/nicknames for the victim.
    - Honey, sweetheart, sweetie, darling, cutie and similar.
- Predators are seemingly open about their age.
    - Stated age creates a potentially predatory conversation compared to victims stated or assumed age.
    - Given age by predators can potentially be fake.
- Predators are in some cases open about them knowing what they are doing is wrong.
- Predators sometimes have a better developed written language.
- Predators use a lot of energy on making sure of secrecy.
- Conversations often turns sexual, in many cases quite early in the conversation.
- Predators often ask a lot of questions, and often more questions than the victims.
    - Sexual questions.
        - Previous sexual experience.
        - If the victim would like the predator to perform different sexual actions to him/her.
        - What sexual actions the victims would like to do or fantasizes about.
    - Questions to ensure no one are monitoring the communication between the predator and the victim.
    - Questions to find out if victims are alone.
    - Ask for pictures.



Conversation
hi
51 tucson dad
hey back
13 tucsion daughtre
kool
sounds fun
ur hot baby girl
lol, ty
look so yummy
ever hook up from here
maybe
mmmmmmmmmmmmmmmmmmmmmmmm
lol
i have b4
yea, but anybody my age
like ur a bit older
yes lots
we have fun
dood, ur cute but married
so i am such a good teacher and make you feeel awesome
you like to be licked? until you orgasm a lot
lol
yes, couple of timez
add me lets chat again
mmmmmmmmmmmmmmmmmm
k
i am so good at that baby
u r
you will love it baby girl
oh yeah?
yes
you dont have to do anything except enjoy it
money were ur mouth is
lol
lets hook up
i am near el con
maybe
kool

Figure 4.5: Example of a predator turning the conversation sexual.

  - About friends and family.
  - About hobbies and interests.
- Predators sometimes avoids answering questions.
- Victims often do respond to sexual questions by asking if the different sexual actions in question hurt.
- Victims often live together with parents, mom or dad, or other family or authority persons.
- Victims often try to hide their communication with predators from their surroundings.

4.4.2 False Positives PAN

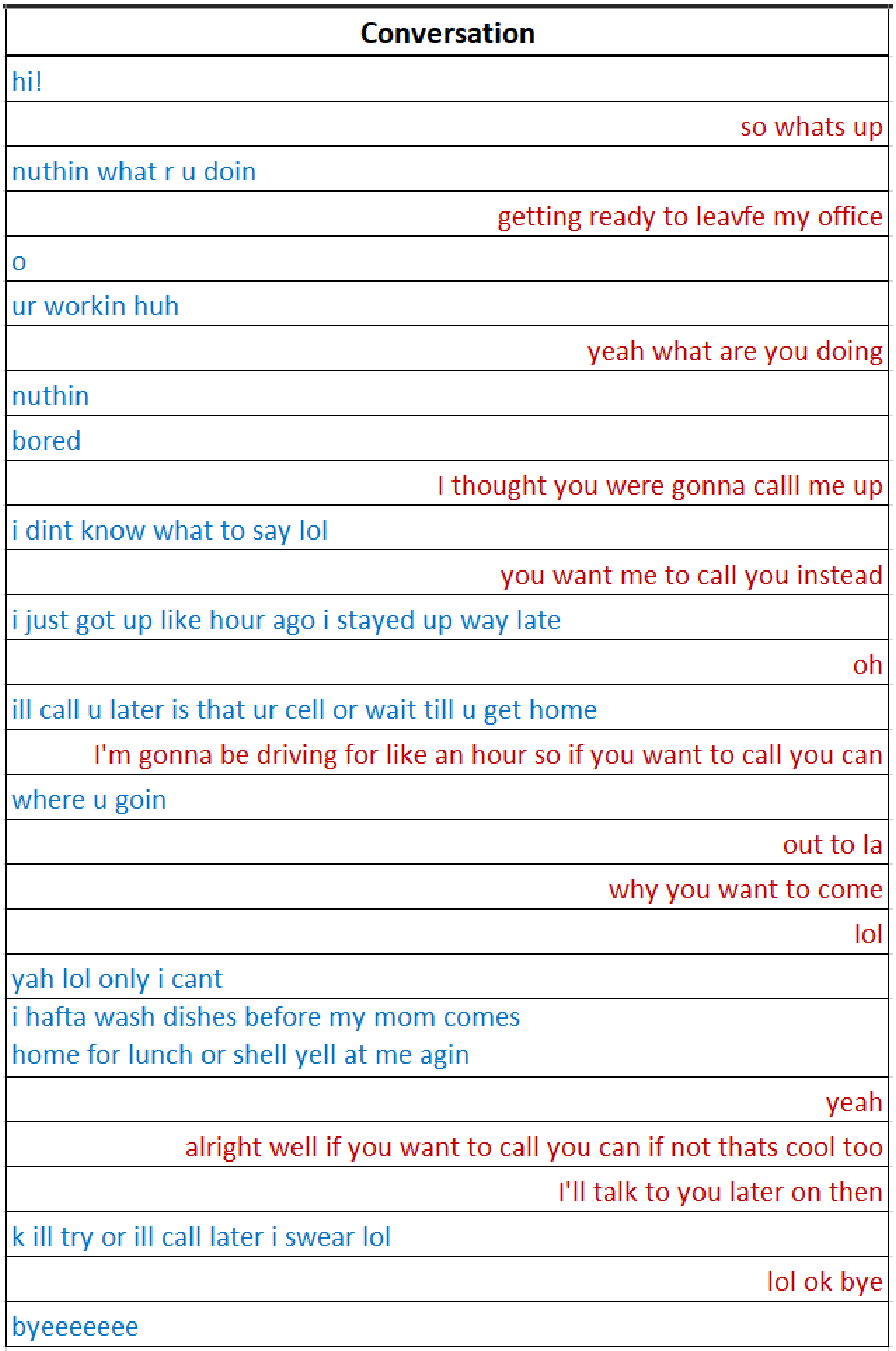
False positive evaluations are the ones evaluated to be predatory which are classified non-predatory by the ground truth.

As with conversations evaluated false positives from the Hybrid dataset, the main reason for the false positive evaluations of conversations from the PAN dataset is also the origin of the dataset with the belonging ground truth making them non-predatory by default. These conversations are coming from chat forums where we do know predators are present, and a lot of these evaluated conversations are also indeed predatory. For example was Omegle utilized as the chat platform and scene of crime when a man in the 40s was convicted for online sexual abuse of a 13 year old minor in Norway during 2021 as stated in [34].

False positives should actually not contain actually potentially predatory conversations. But, because of the datasets origin, we decided it was necessary to highlight predatory features. This because it is highly likely those evaluated false positive actually contain real predators.

Several of the false positives are between two chatters which both are above the Norwegian legal age. As with the false positives from the Hybrid dataset, also evaluations from the PAN dataset where the assumed victim side is just above legal age are of some value. As age of consent varies from country to country [8] and the conversations from the datasets are from all over the world, not limited the borders of Norway, those not considered predatory in Norway could very well be considered predatory in other countries. This means they could still help define how predators work to get their victims.

A very large proportion of the false positive evaluations are sexual conversations. Some of the evaluations are obviously wrong when looking at the conversations, but the evaluation comments points at the fact that they are sexual as a reason for the way they are evaluated. Some of these sexual conversations have no indicators of the two chat parties age, which neither is mentioned in the evaluation comments, and hence the evaluations are solely based on the fact that the conversations are sexual. Such accusations alone are not enough for a conversation to be potentially predatory, and at least not enough to be able to assume it based on only the information which proceeds from the conversation. It is necessary to have more information about illegal matters before a solid conclusion



Conversation
hi!
so whats up
nuthin what r u doin
getting ready to leavfe my office
o
ur workin huh
yeah what are you doing
nuthin
bored
I thought you were gonna calll me up
i dint know what to say lol
you want me to call you instead
i just got up like hour ago i stayed up way late
oh
ill call u later is that ur cell or wait till u get home
I'm gonna be driving for like an hour so if you want to call you can
where u goin
out to la
why you want to come
lol
yah lol only i cant
i hafta wash dishes before my mom comes home for lunch or shell yell at me agin
yeah
alright well if you want to call you can if not thats cool too
I'll talk to you later on then
k ill try or ill call later i swear lol
lol ok bye
byeeeeeee

Figure 4.6: Example of a victim introducing information about living situation.

can be made. However, conversations with only some elements making evaluators react are useful in order to find elements considered potentially predatory from evaluations.

As with the true positives, there are also predators in the false positive evaluated conversations stating they do know what they are doing is wrong and illegal. Also in these cases they seem to seek compassion from the victims and them to say it is nothing to worry about and that they will keep quiet. The admitting of them doing something wrong or illegal often comes after some time, so the predators and victims have had the opportunity to get to know each other first and established a relation. As with everything else, it is much easier to blow off someone you do not know and do not have a relation to, and it becomes much harder when you have a relation to a person. This is also what is highly likely the reason to why such admissions often do not come at an earlier stage of the conversations. This because the predators probably are well aware of the fact that it is very much harder for the victims to cut them off it they do not established some sort of relations before telling it.

Features from conversations evaluated false positive:

- In some cases the chatters seem to know each other from before the conversation in question.
- Many conversations are sexual.
  - Both chatters are above legal age.
  - One chatter is underage and the other is above legal age.
- Predators often say "I love you" to their victims.
- Predatory side uses a lot of nice words/nicknames for the victim.
  - Honey, sweetheart, sweetie, darling, cutie and similar.
- Predators sometimes impersonate as children.
  - One of the chatters have a better developed written language than what to expect from a person of stated age.
- Predators sometimes have a better developed written language.
- Predators sometimes adapt their written language to match the victims language.
- Predator side wants pictures, often nude pictures.
  - Predators in some cases avoids sending pictures back (due to response from victims).
  - Predators in some cases sends fake pictures in return as part of a picture trade.
- Predators early want to establish multiple points of contact by trying to get MSN, Yahoo, phone number, email or other ways of being able to communicate to the victim.
- Predatory side asks a lot of questions, and very often more questions than the victims ask back.

    - Sexually oriented questions.
    - About what the victims are wearing.
    - If the victims would like to do things to please or satisfy the predators.
    - If the victims would like the predators to perform certain specific actions, mostly sexually oriented.
    - If the victims can perform sexual actions on themselves during the conversation.
    - About friends and family.
    - About hobbies and interests.

- Predators are in some cases open about them knowing what they are doing is wrong.
- Predators use a lot of energy making sure of secrecy in order not to get caught.
- Predators often tempt their victims with the ability to buy alcohol, weed, electronics and other things children are not allowed to have or can have a hard time getting on their own.
- Age of victims are not always revealed in chat.
    - Can be estimated based on different information given about the victims.
        - At home with parents.
        - Doing homework.
        - Not allowed to stay up late.
        - Not allowed to go outside in the evening.
        - Grounded.
- Victims often talk about going to school, having to do homework and meeting friends.
- Victims often respond to sexual questions by asking if the different sexual actions in question hurt, whether it is regarding sexual actions for when they meet or sexual actions the predators ask the victims to perform on themselves during the conversations.
- Victims often live at home with parents, mom or dad, or other family or authority persons.

4.4.3 True Negatives PAN

True negative evaluations are the conversations classified actually non-predatory by the ground truth which also are evaluated to be non-predatory.

These conversations are the ones about all and nothing. They are what most people would consider normal in the sense that they contain non-predatory and harmless content between two chatters. Almost all of the evaluation comments states that the true negative conversations are normal conversations. As with the true negative evaluations on the Hybrid dataset, normal is a subjective and somehow challenging definition.

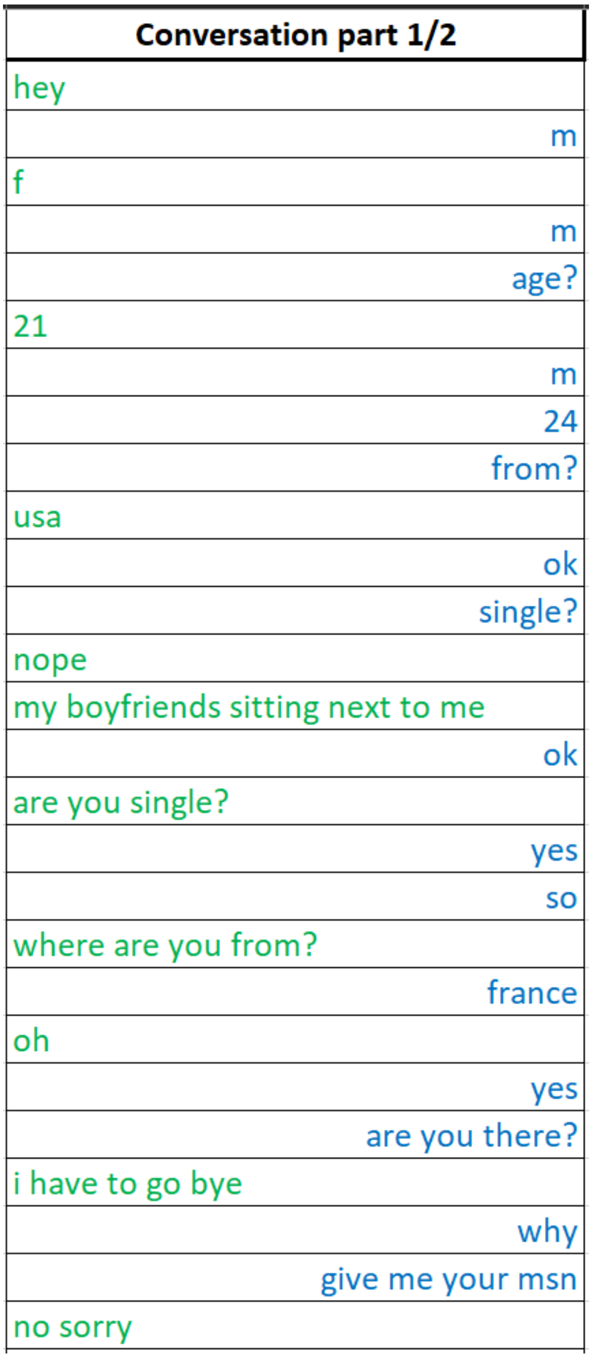
Normal conversations in many cases includes getting to know each other, exchanging basic information about age, place of living and gender. Figure 4.7 is an example of a non-predatory conversations where two people are exchanging such information.

Out of the 3516 true negative evaluations from PAN, only 27 did have a change in decision. This corresponds to 0,77% of the conversations, which is an even lower percentage amount than we saw on the true negatives from the Hybrid dataset. As with the conversations from the Hybrid dataset, this indicates that once a really non-predatory conversation is discovered and found to be non-predatory, the conversation is also actually non-predatory.

The proportion of sexual conversations is 12,80%, making the proportion of sexual conversations quite small and almost half of what it was for the true negatives from Hybrid.

Features from conversations evaluated true negative:

- Normal conversations.
    - Between adults.
    - Between adults and minors.
    - Between minors.
- Most conversations are non-sexual.
- Written language are often of relatively equal level.
- Some conversations are ordinary talks between people trying to get to know each other.
    - Between both young people and adult people.
    - Sharing basic information.
    - Sharing contact details.
    - Discussing hobbies and common interests.
- About technical topics often containing very technically specific terms.
- Adults or people above legal age talking sexual.
    - Cyber sex.
    - Sexual preferences.
    - Sexual wants and needs.
- Conversations are entertaining with funny content.
- Ordinary smalltalk on a wide variety of topics.
    - Sports.
    - School.
    - Work.
    - Hobbies.
    - Interests.



Conversation part 1/2	
hey	
	m
f	
	m
	age?
21	
	m
	24
	from?
usa	
	ok
	single?
nope	
my boyfriends sitting next to me	
	ok
are you single?	
	yes
	so
where are you from?	
	france
oh	
	yes
	are you there?
i have to go bye	
	why
	give me your msn
no sorry	

Figure 4.7: Normal conversation between two chatters exchanging basic information.

4.4.4 False Negatives PAN

False negative evaluations are those evaluated to be non-predatory when they are classified by the ground truth as predatory.

The false negative evaluations from the PAN dataset conversations do in many cases originate from how the dataset was created. Since conversations with breaks of 25 minutes or more were split into separate conversations, there are conversations which are actually parts of longer conversations. This is also proven by looking at all conversations for one given author ID, where some are evaluated true positive and others false negative.

Almost all false negative evaluations are evaluated with the comment that they are normal conversations, i.e. conversations where there are no signs of predatory actions. This makes it hard to gain a lot of meaningful information from the evaluation comments, and the manual analysis is what discovers features of what is normal. As predatory conversations can have been split into multiple conversations due to pause of 25 minutes or more between active communication, some parts of the full conversations can be of just ordinary talk. What is important to remember is the fact that what supposedly is normal in these conversations, is also parts of the preparations performed by the predators. By having normal conversations, they are able to get to know their victims better, get information and potentially gain advantage over the victims. These preparations could e.g. include getting to know the victims better by talking about what the victims likes, family and friends, secrets and all sorts of other things that could further develop relationships. This without any statements of age or signs of age, any sexual urges or planning of meetings. The manual analysis of these conversations also shows that this is the case. Normal conversations are challenging as they are, as the word says, normal. The problem with normal is that normal is also what everyone else are doing, and normal is extremely broad as it in this case includes everything else than what is seen as potentially predatory actions.

What seems to be the most describing features of conversations evaluated false negative is the fact that age is never mentioned and it is also in many cases hard or impossible to assume the age of the two chatters. When age is hard or impossible to tell, the most important factor for judging a conversation to be potentially predatory or non-predatory is gone. As this feature is gone, it becomes much more up to the other features to stand out as something the evaluators would react to, but the false negatives represents the conversations where the evaluators were unable to find substantiating features. The manual analysis found there are a lot of features from these conversations which are similar to features found in other types of evaluations, which makes the age feature stand out as decisive for the evaluation decisions.

Even though most conversations were ordinary talks and normal conversations, some of them were sexual. From the sexual conversations the evaluations did not find any signs of age and therefore did not consider them to be predatory. However, there were 5 conversations which received a change in decision after the whole conversation was disclosed, but only one of those was sexual. Figure 4.8 shows the sexual conversation which first was evaluated to be non-predatory (false negative), but sexual after 18 messages were displayed. When the whole conversation was displayed, the decision was changed as it contains information making it possible to assume the right side of the conversation to be underage. This information reveals that the right side is a girl, probably never have tried any sexual activities, live together with her mom, and is probably underage as she live at home with her mom, goes to school and have no sexual knowledge or experience. The example also shows how the predator drives the conversations forward, especially the sexual part of it. As in other predatory conversations this example also contains a mismatch in questions asked and answered, as the predator asks more questions than the victim which only replies to the questions.

Features from conversations evaluated false negative:

- Chatters seem to know each other from earlier conversations.
  - In many cases the predators state that they have missed the victims.
- Small talk about all sorts of topics to drive the conversations forward.
- Some conversations are about planned meetings.
- It is hard or impossible to assume the age of the two chatters.
- In some cases the predators say things making them sound popular among other people of opposite gender.
  - Predators always follow up such claims by saying they do not care about anyone else then the victims, trying to make the victims feel special.
- Predators often complement the victims with nice words like "you rock", "you're the best", etc.
- Predators use a lot of nice words/nicknames for the victims.
  - Honey, sweetheart, sweetie, darling, cutie and similar.
- Predators sometimes talk themselves down by saying they are boring, not exciting, old, weird, or other things in order to provoke the victims into telling they are not.
- Predators sometimes say things to try come out as cool people, talking about how cool/awesome/fun it is to do things the victims are not allowed to do due to age.
  - Smoke.
  - Drive car.
  - Being allowed to stay up all night.
  - Being allowed to go everywhere.
- Predators tend to seek compassion from their victims in different ways.
  - Telling they are in physical pain, e.g. headache or ill.
  - Telling they are in psychological pain, e.g. have lost someone close to

them or they have been treated bad in a previous relationship.

- Some predators try talking their victims into telling secrets nobody else knows about.
- Predators often tell their victims how nice they will treat them when/if they meet.

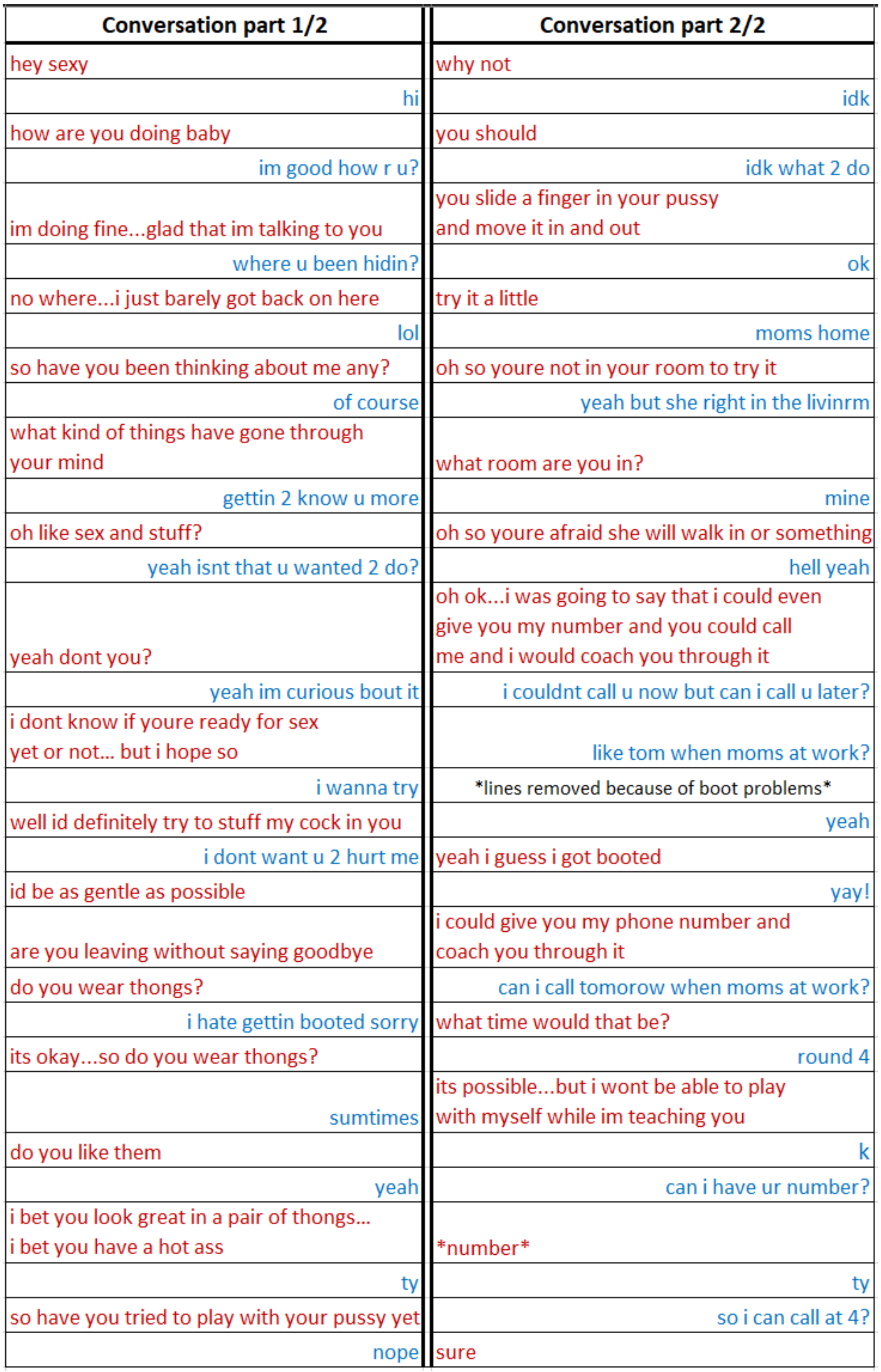
4.5 Most Prominent Features

The analysis has discovered and highlighted a wide variety of different features describing potentially predatory and non-predatory conversations. From this, several are repeating themselves and some are more prominent than others. Based on this, we found the most important features of potentially predatory conversations to be:

- Conversations are sexual.
- Predators use a lot of nice words and nicknames.
- Predators are driving the conversations forward and in general asking more questions than the victims.
- Victims are more passive and reactive, mostly just answering to questions, especially those sexually oriented.
- Predators are often the ones initiating meetings.
- Predators use a lot of energy making sure of secrecy.
- If age is not stated, it is most often possible to assume based on other information given.
  - Living situation.
  - Daily life.
  - Allowed to, not allowed to or have to.

Further, we found the most important features of non-predatory conversations to be:

- Equal curiosity about each other, i.e. balance in number of questions asked and answered.
- Sharing of basic information.
- General talk about various topics chatters care about.
- General talk about everyday life.
- Funny and entertaining content.
- Technical content.



Conversation part 1/2	Conversation part 2/2
hey sexy	why not
hi	idk
how are you doing baby	you should
im good how r u?	idk what 2 do
im doing fine...glad that im talking to you	you slide a finger in your pussy and move it in and out
where u been hidin?	ok
no where...i just barely got back on here	try it a little
lol	moms home
so have you been thinking about me any?	oh so youre not in your room to try it
of course	yeah but she right in the livinrm
what kind of things have gone through your mind	what room are you in?
gettin 2 know u more	mine
oh like sex and stuff?	oh so youre afraid she will walk in or something
yeah isnt that u wanted 2 do?	hell yeah
yeah dont you?	oh ok...i was going to say that i could even give you my number and you could call me and i would coach you through it
yeah im curious bout it	i couldnt call u now but can i call u later?
i dont know if youre ready for sex yet or not... but i hope so	like tom when moms at work?
i wanna try	*lines removed because of boot problems*
well id definitely try to stuff my cock in you	yeah
i dont want u 2 hurt me	yeah i guess i got booted
id be as gentle as possible	yay!
are you leaving without saying goodbye	i could give you my phone number and coach you through it
do you wear thongs?	can i call tomorow when moms at work?
i hate gettin booted sorry	what time would that be?
its okay...so do you wear thongs?	round 4
sumtimes	its possible...but i wont be able to play with myself while im teaching you
do you like them	k
yeah	can i have ur number?
i bet you look great in a pair of thongs... i bet you have a hot ass	*number*
ty	ty
so have you tried to play with your pussy yet	so i can call at 4?
nope	sure

Figure 4.8: Excerpt: Predator leading conversation, turning it sexual and asking more questions. The victim reveals information making it possible to assume age.

Chapter 5

Discussion

This chapter addresses a discussion based on the analysis and results. The discussion aims to consider the differences between features and how they further can be used or not used as parts of hybrid cyber grooming detection systems. The findings from the analysis are also used as basis for different proposed solutions to how cyber grooming detection systems can be improved by introducing knowledge from human analysis making them hybrid systems.

Predatory conversations can occur in all shapes and forms as shown in the analysis. There are, however, certain features that tend to repeat themselves shaping recognizable patterns. Some of the patterns are quite clear, and some are less obvious. It can also potentially be quite effective to detect normal non-predatory conversations, and not only potentially predatory conversations, as it will allow to leave only conversations of potential interest. The main goal with detection is to find the ones that are potentially predatory, but parts of the detection can also be detecting what is not potentially predatory by removing or focusing less on what is non-predatory. This is a challenging task, but the analysis provided good insight into conversation features saying something about what is both predatory and non-predatory.

Some features of conversations are easier to detect than others based on the nature of what they are. These features are also potentially easier to find solutions for how to detect compared to other more complex features which are harder to detect and will require more complex solutions.

5.1 Age

The act of cyber grooming is based on a few premises. One of them is that one chatter of the conversation has to be below the legal age of consent (underage) and the opposite chatter above the legal age of consent. This makes age a vital feature to be present in order to be able to define a conversation to be potentially predatory. Further it is also a prerequisite fundamental in order to get a conviction if a criminal offense takes place. If age was to be completely unknown, it is impossible to determine if a conversation is potentially predatory or not. This simply because the conversation could be between people above the age of consent, between two children or some other combination which is not illegal. Age is sometimes shared in chat and other times it is not. Even though it is shared in chat or not, both will still have their age and potentially predatory age differences can be possible to detect from other features of conversations. In other words, detection before introduction of age is possible, or detection in cases where age is known between the chatters from somewhere else outside the chat is possible.

It is not necessary to be 100% sure of age in order to evaluate and detect a potentially predatory conversation. As far as the purpose of cyber grooming detection goes in this case, the intention is not to detect the predators or convict someone, but rather detect potentially predatory conversations for further follow-up by humans. This means there is some room for inaccuracy in regards of age detection, opening for interpretation and utilization of age descriptive features in the effort towards a hybrid system. That said, the goal is still to detect as accurate as possible.

One complicating factor in regards of age is potential friendships between chatters below and above legal age of consent. Age is, as highlighted, one of the premises needed for a conversation to be potentially predatory. It is still not illegal to be friends, regardless of age, as long as no criminal offenses take place. Even if it is not very normal for adults to be friends with children, it happens from time to time. Due to this, it is highly important to have more features to support a detection of potentially predatory conversation.

5.1.1 Age Given in Chat

Age has shown from the analysis to be something which quite often is shared between the chatters, and it very often happens quite early in conversations in the stage where people are getting to know each other. Both figures 4.1 and 4.5 show examples of predators of different age being open about their age to their victims and introducing it early. The sharing of age goes for both parties of conversations, and it happens in potentially predatory conversations as well as in non-predatory and normal conversations. Figures 4.2 and 4.7 show examples of a normal, non-predatory conversations where both chatters share their age. Because age is such a general feature which shows to be present in all types of conversations, it is not a stand-out feature of potentially predatory conversations, and it is not a defining feature as it alone does not make a conversation predatory. It is, however, essential to get an understanding of age, or at least age range, in order to say a conversation can potentially be predatory.

Even though age is not a predatory defining feature alone, the statement of age in chat makes further detection of potentially predatory conversations easier. As the goal is to detect potentially predatory conversations, the statement of it in chat can make detection easier for a system. When age is stated, it is easier for the system to further focus detection on actions and other features if an potentially predatory age difference is established. For example the AiBA system can take advantage of detected potentially predatory age difference to increase the total risk of the conversation in question. This detection can further be strengthened by other features. On the other hand, the statement of age can also be used to assume the other way that a conversation is possibly non-predatory if both are stated to be above legal age or both underage. It is, however, necessary to consider the fact that stated age does not have to be true. As long as stated age makes the age difference potentially predatory, a system can detect the difference and find it predatory. It is harder if stated age makes for what is seemingly a conversation between two children.

5.1.2 Impersonation

Age is in many cases clearly stated, but, it is not safe science as several evaluations and the analysis discovered potential impersonation. As far as impersonation goes, it is an act making detection of actual age more complicated when one of the chatters states to be of another age than they actually are. It is a previously known fact [13, 35] that impersonation is used as a tactic by predators in order to trick their victims into thinking they are someone else. Atheer Al Suhairy is an example of such, where the 31 year old pretended to be a 25 year old professional model. He used this cover in order to gain the trust and confidence of several victims below 16 years old, which he further misused [35]. Cases like this in combination with the knowledge of impersonation happening makes it more difficult to rely on age stated in chats.

In some cases it is, however, not too important to detect if impersonation takes place. In the example of Al Suhairy [35], it would not be important to discover the fact that he really was 31 years old when stated to be 25. This simply because a conversation between a 25 year old and a person below 16 still would be detected and considered a potentially predatory conversation based on the age difference, below and over legal age of consent. It is not essential do discover cases of impersonation where the faked age still makes the conversation potentially predatory. What is is essential to discover, on the other hand, are conversations where either the victims states to be older in order to appear more attractive to adults or conversations where the predators states to be below legal age of consent to appear more attractive to minors.

Knowing impersonation is a known trick used by predators in combination with the assumed impersonations found by the evaluators and analysis makes it highly likely there are even more cases of impersonation hiding in the datasets. It is also likely that several conversations evaluated one or the other way actually are conversations where one of the chatters pretended to be someone else with success. The reason why some conversations were discovered to potentially contain impersonation was due to way of writing and level of written language. From other predatory conversations and evaluations we do know that predators are able to adapt their way of writing to fit the written language level of their victims. This includes the use of slang words and abbreviations which are popular and commonly used by younger generations. This is a factor making it harder to detect correct age, as a good impersonation potentially will be able to avoid detection by doing everything right. Doing it right in this context then means to state an age, adapt the level of written language and also make sure the content surpasses what can be expected form a person of the impersonated age.

5.1.3 Age Not Stated

Even though the analysis shows there are a lot of conversations without any statements of age, the fact that conversations in the datasets used for the analysis work have been split into multiple conversations have to be considered. In many of the cases where conversations does not have any statement of age, it seems like the chatters knows each other from before, as the example in figure 4.3 shows. By using author ID to find other conversations from the same author, it shows that previous conversations often includes the introduction of age. Due to this fact, statements of age are actually present in more conversations than it appears from the analysis. This makes it a feature occurring more frequent than it looks like by isolated only looking at one and one conversation from the datasets. It can therefore be argued that age is actually introduced in more conversations than it is in conversations of the datasets on which today's cyber grooming detection systems are trained on.

Most of the time age is given at some point in a conversation. However, there are examples of conversations where age is never mentioned. If age is not mentioned, it should be attempted detected by the use of other means. Conversations without age in plain text are more challenging, but it is not impossible to determine roughly age. A lot of information is often given throughout a conversation. In cases where age is not stated, all the other given information is essential in order to be able to detect age. The analysis shows it is possible to determine roughly age by considering such information.

If available information and features does not make it possible to determine age or roughly age range, it can indicate a potentially non-predatory conversation. It is, however, important not to write it off as non-predatory for that reason alone, as the analysis of false negatives shows conversations without any age defining features. This makes it impossible to say with total certainty based on this. In such situations, other elements could be investigated further, for example information from the username or about the user from the platform utilized.

5.1.4 Daily Life

In conversations where the chatters obviously knows each other from before, they often talk about their day. Age has then probably been introduced in previous conversations, and is not stated in chat. In potentially predatory conversations, such talks often include one of the chatters saying he/she was at school, did homework and hung out with friends after school. The other chatter often says he/she went to work, drove somewhere and went to a bar. Information like this makes it possible to form picture of how old the parties are. A person going to school, having to do homework and hanging out with friends after school is highly likely to be of lower age and also most likely underage. An older student would probably not referred to school work as homework, but rather assignments, thesis or similar. When it comes to the other person of the example, driving a car calls for a certain age to be allowed to do. Going to work all day is not something an underage would do, as underage are mostly in school. Going to bars and drinking alcohol is not allowed under certain ages, so in this example it is possible to assume this person most likely is 18 years old or older. The conversation shown in figure 4.3 is an example of such conversation where their ages are not disclosed. It is, however, given enough information to be able to assume the left side of the conversation to probably be below 16 years old and the right side above 18 years old. The information in this example making the left side the potential victim, is that she is going to school and talks about the teacher bossing them around. Further she has homework, and a grandmother (gma) which the predator is wondering if cares about her clothing. The information on the predator, on the other hand, is not as good. But the fact that he is working night and checking email in combination with the way of sexualizing the conversation and asking if the victims grandmother cares, makes it possible to assume that this person most likely is at least above 18 years old.

Information about daily life can therefore be quite valuable as support information. It can either serve to support a stated age or as a part towards detecting age. It can also work the other way if someone does not have their story straight, and serve as part of uncovering impersonation of age.

5.1.5 Living Situation

Another important feature found from the evaluations and analysis is the living situation of the victims. As adult people tends to live by themselves away from parents or other family, the victims of predatory and potentially predatory conversations mostly live together with family or similar. Be it mom and dad, mom or dad, step mom or step dad, grandmother and/or grandfather, aunt, uncle or other family relations or other authority persons. This is an essential feature which tends to come up during conversations, either as the victims have to ask for permission to do something, by question about living situation from the predators, or in other ways where it naturally becomes part of the conversation.

Because the datasets contains split conversations, a lot of the conversations analyzed did have situations where this topic was introduced, and the predators were in most cases the ones to ask in order to find out. In cases where they seemed to know each other from before, the living situation appeared to be already discussed and known. This came to light as for example the predators asked if the parents of the victims were home, either out of the blue or in situations where it would be more appropriate for the victims to be alone, e.g. if sexual actions were proposed to be preformed on web camera. Figure 4.6 is an example of a conversation where the victim says dishes have to be washed before the mother gets home and thus reveals the living situation to be at least together with mother. Even though a person of 16 years old or older could live at home together with mother and have responsibilities in terms of housework, such situation can be used as parts of an age assumption.

Living together with parents or others must not necessarily be a true indicator of underage. As highlighted, a person of 16 years old and upwards can also be living at home before moving out. Another scenario is if parents are getting old, sick or similar. Then it might be necessary for their children to live with them in order to be able to take care of them. It could also be more reasons to why someone would live together with parents as they get older. This makes living situation a less valuable standalone feature. But, it is not without any value, as it can be quite describing in cases where it is combined with other information, for instance not being allowed to go outside in the night.

5.1.6 Allowed, Not Allowed or Have To

As the victims are underage, they are often subject to parental control. By parental control, we mean the parents of the victims, or other authority figures in their lives, are to some extent in charge of their lives in regards of what they are allowed to do, not allowed to do or telling them they have to do something. The analysis found this to be a distinctive feature which often occurred during conversations. This can very often serve as an indicator of both age and living situation. Example of such is if the victims were asked by the predators to go outside in the evening or travel somewhere for the weekend. The response from the victims then were that their parents, or adults in charge, did not allow them to do so. This can be considered a quite unique feature of conversations involving at least one person being underage. It is, however, also a feature which can occur in conversations between two underage people too, as they both then will be subject to their respective authorities. In such situations where the conversation is between two actual underage people, it is less likely that one of the chatters ask the other to go outside to for example meet late in the evening or by night. This because if both chatters are actually underage, chances are they both have rules to follow. Still it is not absolutely certain, as there are differences in families, and some does not have as much boundaries at home as others. This makes it possible to miss conversations where boundaries are absent, but in cases where boundaries are mentioned, it is highly likely the one bringing it up is underage or close to underage.

In cases where the victims are underage, the age itself also sets some restrictions in regards of what they are allowed to and not allowed to. An underage person is for instance not allowed to buy alcohol, go to bars or drive a car. However, this also goes for persons of 16 years and above also. In Norway this is allowed when turned 18, but it varies from country to country. Natural limits set by age is therefore not very useful alone, but can serve as a supporting features in combination with other features.

5.2 Meeting

Predators are often quite eager to meet their victims. This became very clear after the analysis of evaluations and conversations. In some cases they are straight to the point asking to meet within short time after the conversation starts, like in the example of figure 4.3. When meetings are initiated this early in conversations, it is easier to label them as potentially predatory if the age difference defining a predatory conversation is present. In other cases the predators are using more time to establish a trusty relationship before initiating and planning an actual meeting. When more time is used, it is harder detect based on planning of meeting.

Predators use a lot of different approaches when planning or initiating meetings. Sometimes they ask to come over to the victims home, other times they ask to meet on neutral ground. Neutral ground could be for example a hotel. What is interesting to see is that they rarely asks the victims to come to their place. They seem not to be bothered by distance, as it very often is some distance between the victims' and predators' residences. Predators solve this in most cases by offering to come to the victims.

5.3 Attention and Driving Force

In conversations between predators and victims, the analysis shows predators tend to be the more active party. This means predators are the driving force of conversations, being the ones driving them forward. There is not one definition of how conversations are driven forward as each and every conversation is unique. That said, there are a lot of similarities and repeating patterns.

The most prominent trend is that predators use questions to drive the conversations. Predators generally tend to ask a lot more questions than the victims. Figure 4.8 shows an example of a conversation where the predator asks a lot more questions than the victim, and the victim to some extent just answers questions without replying any new questions. This especially goes for the sexually oriented ones. The questions asked by predators are not necessarily the same, but often they are of the investigative type, i.e. questions which most likely will provide informative answers.

Apart from questions, predators also use statements to drive the conversations forward. Most statement driven conversations are also sexually oriented, and the statements are often related to sexual actions and activities the predators would like to perform to the victims.

5.3.1 Nice Words and Nicknames

Everything predators do to their victims are parts of a bigger picture. Small things are done in different situations in order to reach their goal of establishing a solid relation before taking advantage of it. One minor detail, which in itself is nothing but harmless, is excessive use of nice words and nicknames.

The excessive use of nice words and nicknames has shown to be a prominent feature of how predators behave. Exactly why they do it, is not possible to say with certainty, but it is highly likely as a step to try create a feeling of care for the victim. It is a known fact that predators use a wide variety of techniques to establish relations to their victims, and the excessive use of nice words and nicknames can possibly further enhance the feeling of care and comfort for the victims.

The words used varies. Among those that appear to be most in use are words like "love", "honey", "sweetheart", "sweetie", "darling" and "cutie". The words are used in a wide variety of situations. Sometime just a few times during a conversations, when necessary to address the opposite conversation party. In other cases it is used frequently throughout the conversations, and sometimes at the end of nearly every message sent by the predators. Very sensational is the fact that predators in many cases says "I love you" to their victims. This often happens after what has to be considered very short time, after chatting with each other. As the victims are too young to really understand the meaning and power of that sentence, they very often says it back. Sometimes they, however, do not say it before the predators asks if the victims love them back. This can be seen as a strong indicator of a potentially predatory conversations, as the predators more or less are "forcing" the victims to say it back.

Non-predatory conversations tend not to have excessive use of nice words and nicknames. It happens from time to time that nice words or nicknames are used, but not in the same way as in predatory conversations. It is still necessary to consider where the conversations come from, as conversations from other sources probably would contain such features also for non-predatory conversations. For example will it be very natural for a chat between an adult love couple to include excessive use of nice words and nicknames, and say "I love you" to each other. Even though this is not found from this analysis, it is something to take into consideration not making it an absolute feature of detection.

5.4 Sexual Conversations

Sexual conversations show to occur on a quite frequent basis in the datasets used for this thesis. Of the true negative evaluations from the Hybrid dataset, 20% were marked as sexual, and 12,80% were marked as sexual from the PAN dataset. Since the number of conversations evaluated from the PAN dataset was significant larger than from Hybrid, the percentage from the PAN dataset is probably of better accuracy than the percentage from the Hybrid dataset. Anyways, the proportion of sexual conversations occurring from a set of non-predatory conversations is in-between 12,80%-20%, which makes it a fairly large proportion. Due to this, the fact that a conversation is sexual alone is not enough to have it labeled as potentially predatory. Given the percentage numbers found for the non-predatory conversations, these numbers are probably not representative in general for all conversations. The conversations gathered for the PAN-2012 dataset originates from sources where it could be possible that sexual conversations occurs on a more frequent basis than it does on other chat platforms. The percentage in itself is, however, not that important. The important take-away from it is the fact there are sexual conversations among non-predatory conversations. As of this, sexual conversations alone are not enough for conversations to be predatory.

A lot of the conversations identified as predatory from the data collection experiment were identified as predatory only because they were sexual. It applies to conversations identified both true positive and false positive. For the true positive evaluations, this is correct, but sexual as a standalone feature is not enough. For the false positives on the other hand, some of them showed to be only sexual, but between adults, making them non-predatory as they also are classified. This can indicate that ordinary people tend to emphasize sexual content to a very large extent when evaluating conversations to be predatory. This is absolutely a major feature of predatory conversations, and also probably one of the most significant ones. It is still necessary to know more about the chatters than just the fact that the conversation is sexual.

Sexual conversations did occur on a quite frequent basis between adults in the datasets used for this thesis. As the datasets originates from, among other things, platforms where it is possible to meet random people, it is possible that adults seeks there to engage in cyber sex for excitement. What differentiates sexual conversations between adults is that chatters tend to be equally active in regards of driving the conversation forward. This means they are both contributing almost equally to the conversation.

In some cases separating sexual conversations between two adults and conversations between one adult and one child can be challenging. Even though most sexual conversations between adults and children can be detected based on the children being more reactive to sexual questions, there are also some sexual conversations where the children are more active and obviously more interested in sexual conversations. Such conversations can make detection based on sexual content harder, but it is still a good feature for discovery of situations where the victims seems to have no experience and fear in regards of if it hurts or not.

One prominent feature of sexual predatory conversations is that victims very often have little or no previous knowledge or experience on the area. Predators on the other hand, seem to have extensive knowledge and experience. The predators are turning the conversations sexual and the ones driving them forward. In general predators tend to be the more active party of conversations being the ones to drive them forward and ask the most questions. When predators are driving sexual conversations forward, the victim often becomes even more passive and mostly reactive to questions. Meaning the victims mostly just answers to the questions asked, and the answers are very often simple and confirming, denying or accepting, e.g. various variations of yes, no and ok.

In situations where age is not known, a revealing factor is often how victims responds to sexual questions. As they tend to have little or no previous experience, they are often unsure whether different sexual actions hurts or not, and how things work. This emerges in cases where they ask questions back. These questions are often straight forward, asking if it will hurt, or stating they will not participate in something that would potentially hurt them. In some cases they are also asked by the predators to perform sexual actions to themselves while chatting. The victims then often ask how to do it. The predators then instruct them either by chat or offer to call and coach the victims through it while talking.

5.5 Secrecy

Predators are doing something wrong when they initiates predatory conversations. From the analysis it is obvious that most of them are well aware of what they are doing, and that they are doing something wrong. Common to all of them is that they seem to not care. In some conversations predators even states the fact what they are doing is illegal, as in the example of figure 4.4. In many cases it appears like they are saying it just to get the victims approval to keep the communications going and help hiding it. Help hiding it means the victims makes an extra effort to keep their communication a secret and hidden, at least from the authority figures in the victims lives. In some cases the predators actually wonder if the victims are to tell their friends about their relation, but this tend to be rather unusual.

If adults are communicating with children without any other intentions than having nice friendships, there is no need to try hide the communication. In such cases the adults probably do not think they are doing something wrong either, and therefore naturally make no effort into secrecy at all.

From the analysis it was discovered that predators use a lot of energy making sure of secrecy. By making sure of secrecy, we mean predators put a lot of effort into making sure no one is aware of the relation between the predators and victims. This is often done by asking a lot of questions and sometimes making statements. The types of questions used are often of the investigative type, seeking to find out where the victims are sitting when communicating, if there are other people around able to see their screen, if someone potentially could walk in on them and similar questions. Depending on the answers given by the victims, the predators sometimes also state, in various ways, it is important that they hide the their communication.

5.6 Normal Conversations

Normal conversations, i.e. conversations which contains no potentially predatory traces, are the most frequent types of conversations. They consists of content considered as normal by most people. Everyone has an opinion of what is normal, but normal is hard to describe and generalize.

From the analysis we found that conversations evaluated to be normal for this study very often included curiosity about each other in order to get to know each other. Due to this, the normal conversations most often consisted of the chatters asking questions back and forth in order to get to know each other. Basic information about place of living, age and gender was typically shared very early. Further, the conversations often turns towards hobbies, interests, school, work and other everyday conversation topics, or funny and entertaining topics.

Normal conversations can be between chatters of all ages, two children, two adults, or adult and child. In cases where it is between two equal parts, e.g. two adults, the analysis found the level of their written language to often be quite equal, whereas there sometimes are mismatch in level of written language in predatory conversations.

The analysis also found a whole lot of technical conversations. Technical conversations contains professional IT language and are strictly professional and normal. Such conversations can easily be discovered due to the excessive use of technical terms and abbreviations. It is, however, essential to note that these technical conversations are present in the datasets used due to how the PAN-2012 dataset was constructed. Even if such conversations are present to some extent in the datasets used for this thesis, it is likely to believe the amount of equivalent technical conversations are not present to an equal degree for regular chat providers. It is also conceivable that potentially predatory conversations can be somehow technical once in a while, as for example if a predator is telling a victim how to install a program to circumvent cyber grooming detection or anything else.

The most challenging part of normal is that it happens in between in predatory conversations as well. Several of the false negative evaluations stated the conversations to be normal, containing nothing potentially predatory. This means that normal also goes as part of predatory preparations which makes it not exclusive for non-predatory conversations.

5.7 Consistency in Features

As the analysis and discussion have shown, conversations consist of multiple features. Conversation features can be used a long way to either evaluate a conversation to be non-predatory or predatory. The more prominent they are, the easier it is to use them.

There is a lot from human analysis that can be utilized in a hybrid detection system. It is still necessary to consider the fact that most features that could be seen as very valuable in order to detect potentially predatory conversations also often are not absolutely consistent. This comes as a result of every conversation being unique. Even though there are a lot of similarities, there are always some differences, large or small, which makes it impossible to find features being absolutely consistent. Because of this, one feature alone cannot be accounted for as the one and only in order to detect a potentially predatory conversation. The goal is, however, not to find one standalone feature to detect them all. The goal is rather to find multiple features, and further allow the combination of them to be the basis for solid detection.

No features are absolutely consistent, but by considering several features at the same time, it is highly likely one or more of them are present in predatory conversations. The different features discussed have shown to quite often be present in conversations. Some of them have from time to time been missing, but in these cases several of the other features have been present. It is almost never just one, it is mostly several. Meaning, even though one specific feature is not present in all conversations, at least two or more various features are mostly present at all time. This is an advantage for detection of potentially predatory conversations, as it can be exploited in order to detect the potentially predatory conversations.

It is also important to note even if some features are present in a conversation, it is not always absolute. This meaning for example if a chatter is telling he/she lives with the parents, it could be the parents are old and need care from their adult child. In such situation, the feature of living with parents is not an indicator of a potentially predatory conversation.

5.8 How to Make a System Hybrid

This thesis has discovered a wide variety of different features from human analysis which in different ways can be utilized towards the making of hybrid cyber grooming detection systems. The already existing AiBA system [3], based on machine learning classification of text only, can probably benefit from implementing them in order to potentially get even quicker detection of potentially predatory conversations. The system uses a total risk score for the whole conversation which is based on individual risk score for each and every message sent. By implementing features from human analysis, the output from these features can add to the total risk score of the conversation in order to reach an even more accurate development of total risk. The severity of the different features should be considered in regards of how the risk level should change to it. In other words, some features should be weighted more than others, and different combinations of features should be weighted differently. For example, the risk should be increased more (weighted more) in case of imbalance in questions asked and answered, than it should for detecting the chatters age (weighted less). This because age is not a predatory defining feature in itself, but imbalance is to a greater extent.

It could also potentially be beneficial for some of the features to be allowed to raise warnings on their own or in combination with others. For example in cases where the imbalance in questions asked and answered is found to be significant and the age of the chatters is detected to be e.g. 13 and 40.

No features have shown absolute presence in predatory conversations. The key in regards of the features, is therefore to use them in a combined effort. Meaning one single feature alone will never be good enough for the detection of potentially predatory conversations, but combining multiple features will increase the likability of fast and accurate detection.

By utilizing machine learning and Natural Language Processing (NLP) in different ways, the different features can be added to create extra value for the system. The different features are describing in different ways, which is something that can be taken advantage of. As a predatory conversation needs at least two different elements present in order to be defined as potentially predatory, variables can be created and stored for each of them as they occur. These elements can be addressed as:

- What are they?
- What actions do they perform or plan?

What they are is the question of age, child or adult, i.e. if the chatters are above or below legal age of consent. Next, the actions they perform defines if there are any potentially predatory actions or planned actions present in the conversation. By adding to these variables as features occur, the presence of both variables can call for an alarm if they are within the definition area of potentially predatory.

5.8.1 Detection Based on Questions Asked

As predators tend to ask more questions than victims, an implementation counting questions could be added to the system using NLP and Bag of Words (BoW). Bag of Words will allow to keep word count. By counting the number of questions asked and answered by the two chatters of a conversation, it is possible to detect any imbalance. As the imbalance increases, the risk should increase exponentially. If the imbalance is equalized, the risk should decrease. Since predators tend to be the ones asking the most questions and victims the ones to answer the most, such imbalance can be used in systems like AiBA to raise the total risk score of the conversation to the existing risk score development.

5.8.2 Age Detection

Age is, as discussed, a vital part of a potentially predatory conversation. It is often exchanged in chat and often very early in conversations. In other cases age is not stated in chat, but it is possible to determine roughly age based on information given.

When age is stated, it can easily be detected using text based classification detection from machine learning and NLP. Such text classification can be rule-based where a set of predefined linguistic rules and words makes the basis for the detection as words are classified into different defined groups.

For cases where the age is not stated in the chat, it will be necessary to detect potential age based on other features. For detection in such situations, features like living situation, daily life, allowed to, not allowed to or have to can be utilized in order to create a system implementation based on NLP utilizing Bag of Words (BoW)

5.8.3 Sexual Conversation Detection

Sexual conversations are not potentially predatory defining alone, but in combination with other features it can be a good indicator. For example in cases where an adult is initiating a sexual conversation with a child and asking a lot more questions than the child.

In order to detect a sexual conversation, NLP can be used with for example FastText or lists of words where potential alternative spellings are included. The advantage of using word embedding and FastText is the ability for better representation of new, rare or misspelled words. As predators sometimes tend to adapt their language to fit their victims language, it is likely known words can be used in a misspelled way. The use of FastText will therefore be the better option for detection of such and still get the benefit of better representation of misspelled or rare words.

5.8.4 Normal Conversation Detection

Normal, non-predatory conversations are the most common ones and occur on the most frequent basis. Because potentially predatory conversations can seemingly start off as normal, it is not possible to write off conversations from being harmful because they are seemingly harmless. What normal conversations can be used for, on the other hand, is to lower the risk score of the conversation. The easiest implementation for normal conversations would probably be by machine learning classification, training the model to recognize them. It should, however, not be trained on the data originating from PAN-2012, as this data evidently contains potentially predatory conversations as found from the analysis.

Technical topics and terms have from the analysis proven to be present exclusively in non-predatory conversations, in conversations considered as normal or technical. This can be utilized and taken advantage of, as no predatory conversations have shown to contain it. They can be discovered by the extensive use of technical terms and abbreviations, which to a great extent can be the basis for detection. Machine learning classification will most likely be the better option for the implementation of this. The algorithm can then be trained on datasets including technical work chats and papers, which will make a good training bases for such model.

5.8.5 Feature Summary

By implementing features from human analysis for detection of potentially predatory conversations, the total risk of conversations will have to calculated in another way than today. The different features will add to the risk based on the detection of them. In order for the humans following up the conversations detected to understand why they are labeled as potentially predatory, it could be advantageous to show a summary of features adding up to the score or the detection. It is likely the humans following up does not think of all elements of a conversation, which makes for great value as it can avoid predatory conversations from being ignored due to human error.

5.9 Disadvantages and Experiences

From the work with this thesis we found some disadvantages with the datasets used and the data collection experiment.

5.9.1 Dataset

It could be argued that the datasets of conversations used for this thesis are not ideal. The reason for this is because of two important factors:

1. Predatory conversations do not contain real victims.
2. Some of the conversations classified as non-predatory are most likely actually predatory.

Because the victims of the predatory conversations were posing as victims and not real victims, the features found from the predatory conversations are not organic. This does not necessarily mean they are of less value, but it would be more ideal to use data containing real victims in order to get the best features possible.

The other disadvantage of the datasets is that some of the classified non-predatory conversations are most likely actually predatory conversations. This is an even bigger problem when used for classification in machine learning, but it did also create imbalance in the evaluations for this thesis.

5.9.2 Data Collection Experiment

The data collection experiment was much harder to perform than initially thought. The task itself did not seem to overwhelming, but it did require some more effort from the participants than e.g. an ordinary survey would. It was, however, hard to convince people into participating, making the initial number of participants rather low. This number should ideally have been bigger. Next, almost half of the initial participants did not do any evaluations, leaving the actual number of participating participants even smaller.

The different participants did not do an equal amount of evaluations. Some did just a few conversations whereas others did dozens. The total number of conversations evaluated was not bad, but it should preferably have been bigger and contained evaluations from more different people.

The dataset used for the data collection experiment was too big in regards of number of conversations. Due to this, not many conversations got more than one single evaluation. Ideally each conversation should have been evaluated multiple times, as this would have allowed to do more analysis work on each conversation with corresponding evaluations. For this situation it would probably have been better to have a selection of around 100 conversations to be evaluated in order to get multiple evaluations of each conversation.

If the dataset were to be of the same size for a new data collection experiment, it would probably have given better yield to have more motivated participants. This can probably be achieved by using services like Amazon Mechanical Turk where participants are payed to participate and motivated by money. This obviously comes at a certain cost, but it would probably be worth it in order to get more evaluations and several evaluations of each conversations, coming out as the best solution if the experiment was to be performed one more time.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

This thesis has investigated if features extracted from human analysis can be used for improving a cyber grooming detection system based on a machine learning algorithm. Features from human analysis were gathered in three different ways which resulted in the analysis and results, and discussion for this thesis. A data collection experiment was conducted to get evaluations from ordinary people. Evaluations performed by a summer intern were retrieved, and content analysis of all the gathered evaluations and conversations was performed. The evaluations performed by participants in the experiment and the summer intern contained a lot of useful information about what features of conversations humans found to be descriptive of either predatory or non-predatory conversations. These evaluations and the corresponding conversations were further analyzed for the discovery of trends and patterns.

From the analysis, several prominent features were found and elaborated on. No predatory features stood out as absolutely consistent in all conversations, which was an interesting discovery. Even though there was not one significant feature present all the time, several of the features were found to almost always be present in different combinations. In other words, if one feature was not in the conversation, minimum two others mostly were.

Predatory conversations showed to contain various describing features which in different ways can be used to extract information about conversations. Features like living situation, daily life and things persons are allowed to do, not to do, or have to do can to a great extent say something about age or age range. Sexual conversations and the desire for secrecy in combination with meeting initiation can often tell something about the predators motives and intentions. Imbalance in questions asked and answered between the two chatters can often indicate a predatory conversation and a predatory age difference, where the victim often is the more passive and reactive chatter.

Non-predatory conversations showed to be mostly normal conversations, but some of them were also sexual. Normal is very broad, but from the analysis we were able to generalize it to some extent. Normal in this context are conversations where the chatters are sharing basic information about themselves, talking about everyday life, and funny and entertaining conversations. The challenging part about normal, however, is that it also can occur in predatory conversations. Non-predatory conversations differentiates from predatory conversations as the chatters tend to be more equally curious about each other, making a better balance in questions asked and answered between the chatters. One, almost absolute, feature separating non-predatory conversations from predatory conversations is when conversation are technical. In such situations, the analysis found the conversations to be exclusively non-predatory, but this must be considered probably is due to the dataset they originate from. It is, however, conceivable for potentially predatory conversations to potentially be technical from time to time if the predators are instructing victims to for example install software to avoid cyber grooming detection.

The knowledge of several features being present in potentially predatory conversations calls for beneficial opportunities in regards of detection. A system can be built in a way where features are discovered individually and adding up to the total conversation risk score. This can further potentially allow for faster detection due to the severity of the detected features. Combinations of certain types of features can also be used for raising warnings outside the total risk evaluation for the conversation. Various machine learning methods can be utilized for the purpose of implementing detection of features, such as classification and Natural Language Processing (NLP).

6.2 Future Work

For future work, the proposed implementations can be implemented into AiBA or other cyber grooming detection systems. The proposed implementations can potentially improve the detection capabilities of systems and allow for earlier and better detection. The thesis has highlighted several other features which could be looked more into how to implement in cyber grooming detection systems.

To be able to get an even better basis of data, a more flawless dataset can be created in order to get more accurate training of machine learning models, and also for potentially further collection of evaluations. Such evaluations should preferably be performed by motivated participants in order to reach a sufficient number of evaluations, which will allow for extended analysis of evaluations. This can be done by creating a selection of conversations which is desirable to get evaluated to constitute the dataset, which further is used to get evaluations from Amazon Mechanical Turk.

Bibliography

[1] K. Kopecky, "Cyber grooming danger of cyberspace," *Olomouc: Net University*, 2010.

[2] J. G. Noll, L. A. Horowitz, G. A. Bonanno, P. K. Trickett, and F. W. Putnam, "Revictimization and self-harm in females who experienced childhood sexual abuse: Results from a prospective study," *Journal of Interpersonal Violence*, vol. 18, no. 12, pp. 1452–1471, 2003.

[3] P. Bours and H. Kulsrud, "Detection of cyber grooming in online conversation," in *2019 IEEE International Workshop on Information Forensics and Security (WIFS)*, IEEE, 2019, pp. 1–6.

[4] D. Michalopoulos, I. Mavridis, and M. Jankovic, "Gars: Real-time system for identification, assessment and control of cyber grooming attacks," *Computers & security*, vol. 42, pp. 177–190, 2014.

[5] S. Craven, S. Brown, and E. Gilchrist, "Current responses to sexual grooming: Implication for prevention," *The Howard Journal of Criminal Justice*, vol. 46, no. 1, pp. 60–71, 2007.

[6] S. Craven, S. Brown, and E. Gilchrist, "Sexual grooming of children: Review of literature and theoretical considerations," *Journal of sexual aggression*, vol. 12, no. 3, pp. 287–299, 2006.

[7] *Lov om straff (straffeloven) - Kapittel 26. Seksuallovbrudd - Lovdata*, [Online; accessed 4. Nov. 2020], Nov. 2020. [Online]. Available: `https://lovdata.no/dokument/NL/lov/2005-05-20-28/KAPITTEL_2-11#%C2%A7291`.

[8] AgeOfConsent.net, *Legal ages of consent by country*, [Online; accessed 24. Nov. 2021], Nov. 2021. [Online]. Available: `https://www.ageofconsent.net/world`.

[9] J. Taylor, "Online investigations: Protection for child victims by raising awareness," in *ERA Forum*, Springer, vol. 16, 2015, pp. 349–358.

[10] T. Charman, A. Campbell, and L. S. Edwards, "Theory of mind performance in children, adolescents, and adults with a mental handicap," *Cognitive development*, vol. 13, no. 3, pp. 307–322, 1998.

[11] E. Commission, *Fight against child sexual abuse - Migration and Home Affairs - European Commission*, [Online; accessed 14. Nov. 2020], 2016. [Online]. Available: `https://ec.europa.eu/home-affairs/what-we-do/policies/cybercrime/child-sexual-abuse_en`.

[12] D. Holloway, L. Green, and S. Livingstone, "Zero to eight: Young children and their internet use," *LSE, London: EU Kids Online*, 2013. [Online]. Available: `https://ro.ecu.edu.au/cgi/viewcontent.cgi?article=1930&context=ecuworks2013`.

[13] M. Ashcroft, L. Kaati, and M. Meyer, "A step towards detecting online grooming–identifying adults pretending to be children," in *2015 European Intelligence and Security Informatics Conference*, IEEE, 2015, pp. 98–104.

[14] R. Williams, I. A. Elliott, and A. R. Beech, "Identifying sexual grooming themes used by internet sex offenders," *Deviant Behavior*, vol. 34, no. 2, pp. 135–152, 2013.

[15] S. Theodoridis, *Machine learning: a Bayesian and optimization perspective*. Academic press, 2015.

[16] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of machine learning*. MIT press, 2018.

[17] I. Kononenko and M. Kukar, *Machine Learning and Data Mining: Introduction to Principles and Algorithms*. Horwood Publishing Limited, 2007.

[18] J. Eisenstein, *Introduction to natural language processing*. MIT press, 2019.

[19] H. Lane, C. Howard, and H. Max Hapke, *Natural Language Processing in Action: Understanding, analyzing, and generating text with Python*. Manning Publications Co., 2019.

[20] K. Chowdhary, "Natural language processing," in *Fundamentals of Artificial Intelligence*, Springer, 2020, pp. 603–649.

[21] C. Manning and H. Schutze, *Foundations of statistical natural language processing*. MIT press, 1999.

[22] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," in *European conference on machine learning*, Springer, 1998, pp. 137–142.

[23] H. (Kung-Hsiang, "Word2Vec and FastText Word Embedding with Gensim - Towards Data Science," *Medium*, Feb. 2020. [Online]. Available: `https://towardsdatascience.com/word-embedding-with-word2vec-and-fasttext-a209c1d3e12c`.

[24] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

[25] T. Ganegedara, *Natural Language Processing with TensorFlow: Teach language to machines using Python's deep learning library*. Packt Publishing Ltd, 2018.

[26] F. Muñoz, G. Isaza, and L. Castillo, “Smartsec4cop: Smart cyber-grooming detection using natural language processing and convolutional neural networks,” in *International Symposium on Distributed Computing and Artificial Intelligence*, Springer, 2020, pp. 11–20.

[27] P. Bruce, A. Bruce, and P. Gedeck, *Practical Statistics for Data Scientists: 50+ Essential Concepts Using R and Python*. O’Reilly Media, 2020.

[28] L. Derczynski, “Complementarity, f-score, and nlp evaluation,” in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16)*, 2016, pp. 261–266.

[29] C. J. Van Rijsbergen, “Foundation of evaluation,” *Journal of documentation*, 1974.

[30] P. Justice, *Perverted-Justice.com - The largest and best anti-predator organization online*, [Online; accessed 22. Nov. 2020], Nov. 2020. [Online]. Available: `http://www.perverted-justice.com`.

[31] G. Inches and F. Crestani, “Overview of the international sexual predator identification competition at pan-2012.,” in *CLEF (Online working notes/labs/workshop)*, vol. 30, 2012.

[32] E. Villatoro-Tello, A. Juárez-González, H. J. Escalante, M. Montes-y-Gómez, and L. V. Pineda, “A two-step approach for effective detection of misbehaving users in chats.,” in *CLEF (Online Working Notes/Labs/Workshop)*, vol. 1178, 2012.

[33] P. D. Leedy and J. E. Ormrod, *Practical Research: Planning and Design*, 11th. Pearson Education Limited, 2015.

[34] P. Solberg and Adresseavisen, *Like etter denne chatten gjennomfører trønderen voldtekten av tenåringsgutten*, [Online; accessed 24. Nov. 2021], Nov. 2021. [Online]. Available: `https://www.adressa.no/pluss/nyheter/2021/11/22/Like-etter-denne-chatten-gjennomf%C3%B8rer-tr%C3%B8nderen-voldtekten-av-ten%C3%A5ringsgutten-24824668.ece?rs4387491637768222938&t=1`.

[35] D. T. L. Hui, C. W. Xin, and M. Khader, “Understanding the behavioral aspects of cyber sexual grooming: Implications for law enforcement,” *International Journal of Police Science & Management*, vol. 17, no. 1, pp. 40–49, 2015.

NTNU

Norwegian University of
Science and Technology