TALLINN UNIVERSITY OF TECHNOLOGY
School of Information Technologies

Nani Khantadze 184046IVSB

Detecting Online Sexual Predatory Conversations in Minors

Bachelor's thesis

Supervisor: Alejandro Guerra Manzanares
MSc in Cyber Security

Tallinn 2022

TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond

Nani Khantadze 184046IVSB

Alaealiste kuritahtlike seksuaalteemaliste internetivestluste tuvastamine

bakalaureusetöö

Juhendaja: Alejandro Guerra Manzanares
MSc in Cyber Security

Tallinn 2022

Author’s declaration of originality

I hereby certify that I am the sole author of this thesis. All the used materials, references to the literature and the work of others have been referred to. This thesis has not been presented for examination anywhere else.

Author: Nani Khantadze

16.06.2022

Abstract

Online sexual abuse featuring in young children became major issue after internet became massively available to people. Vast number of juveniles have been victims of sextortion and yet, has not been found efficient way to tackle the issue.

The aim of this thesis is to evaluate the importance of above-mentioned issue, examine already conducted research papers in the field and try to demonstrate the alternative approach by applying one of the deep learning algorithms, specifically, Convolutional Neural Networks (CNN) algorithm using character based semantic analysis.

Research and implementation were conducted in stages: data collection, data pre-processing, data analysis and presenting the results. The data used for testing purposes is taken from International Sexual Predator Identification Competition at PAN-2012. Initial idea of PAN-2012 competition was to demonstrate how machine learning and text mining techniques can be a huge help for forensic investigators.

This thesis is written in English and is 37 pages long, including 6 chapters, 15 figures and 1 table.

Annotatsioon

ALAEALISTE KURITAHTLIKE SEKSUAALTEEMALISTE INTERNETIVESTLUSTE TUVASTAMINE

Võrgupõhine alaealiste seksuaalne ahistamine on saanud internetiteenuste laialdase leviku ajal oluliseks probleemiks.

Tohutu hulk alaealisi on olnud seksuaalse ahistamise ohvrid ning praeguseks ei ole leitud efektiivset meetodit selle probleemi lahendamiseks.

Käesoleva bakalaureusetöö eesmärgiks on hinnata eeltoodud probleemi tõsidust, läbi analüüsida sel teemal avaldatud teadustööd ja demonstreerida võimalikku alternatiivset lähenemisviisi, rakendades süvaõppe algoritmi, mis põhineb konvolutsioonilistel närvivõrkudel (CNN) ning karakteripõhisel semantilisel analüüsil.

Uuring ja juurutamine viidi läbi etapiviisiliselt: andmete kogumine, andmete eeltöötlus, andmeanalüüs ja tulemuste esitlus.

Testandmed on saadud võistluselt "Rahvusvaheline seksuaalkurjategijate tuvastamise võistlus PAN-2012".

PAN-2012 võistluse algseks ideeks oli näidata, kuidas masinõpe ja tekstikaeve tehnika võivad tohutult aidata kohtu-uurijaid.

Käesolev töö on kirjutatud inglise keeles 37 leheküljel, sisaldab 6 peatükki, 15 joonist ning 1 tabelit.

List of abbreviations and terms

SVM	Support Vector Machine
RNN	Recurrent Neural Network
CNN	Convolutional Neural Network
IWF	Internet Watch Foundation
CADS	Corpus Assisted Discourse Studies
FBI	Federal Bureau of Investigation
XML	Extensible Markup Language
NLP	Natural Language Processing
OCR	Optical Character Recognition

Table of contents

List of figures

List of tables

1 Introduction

The sexual abuse of juveniles has been subsisted in the history of humankind already for a long time. In old days, it could have been drawings, literature or any other way that was available back then. However, in modern era, the accessibility to Internet and social media has dramatically changed the ways of children abuse. Statistics from early 2000s show that child pornography cases increased from 22% in 1994 to 69% in 2006 [1]. It has become a challenge for law enforcement to control the methods of sex offenders.

Mainstream online platforms for sexual offenders to reach children are social media, chat rooms and online gaming. Federal Bureau of Investigation (FBI) keeps track of statistics about online predators and their actions. Federal investigators believe there are more than 500,000 online predators active each day, having multiple online profiles. Majority of the victims are in between 12 to 15 years old and 89% of victims are contacted via chat rooms and instant messaging [2]. Additionally, research has shown that many of the convicted adults have approaching many children at the same time. The Internet Watch Foundation (IWF) has released annual report stating that in 2019, almost 9 in 10 URLs containing child sexual abuse material were hosted in Europe. For instance, The Netherlands hosts 71% of child sexual abuse content [3].

For the last two years Covid-19 pandemic and confinement orders made people much more dependent and engaged with internet than ever before. Tiktok, Instagram, Facebook became major platforms for youngsters to interact with each other. Massive rise in usage of these platforms have given much more accessibility to online abusers to reach desired audience

Identifying sexual predatory chats manually is an immense work and requires a lot of manpower. Technical approach through machine learning/deep learning helps to tackle the problem in an efficient way. Analysing millions of chat logs to identify potential risk in a way that saves time is crucial for detecting criminals and save victims while running the investigation. Even though, one might try to use technical approach there is a problem of obtaining relevant data, due to the regulatory issues and data sensitivity of it. Most of

the data retrieved so far, is for investigation purposes under the police authorities' hands and is not publicly available.

A *Sexual predator* is defined as a person or group that ruthlessly exploits others [4]. While many research papers and resources use it as to describe the person who obtains or is trying to obtain sexual contact with another person in a metaphorically manner. Also, sexual offenders are used in the same context. This paper will refer these terms (sexual predator, sex offender, predator) interchangeably.

2 Background Information

2.1 Machine Learning

Machine learning gives us ability to learn things about the world from large amount of data, that we as a human being can't possibly study. In the era of big data, having tools that help pre-process data and make predictions based on it became crucial. Teach machines means to learn patterns from looking of examples in data, such that it can later recognize those patterns and apply them to new things that it hasn't seen before.

More scientific definition of machine learning is "A computer is said to learn from Experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E [5]. A well-defined learning problem requires a well-specified task.

Machine learning is divided into three main types: supervised and unsupervised where one is predictive and another descriptive, respectively and reinforcement learning.

Supervised learning goal is to learn a mapping from input x to output y, given a labelled set of input-output pairs [5]. Learning process, we give algorithm already labelled data, with different classes and we expect from the algorithm certain predictions, so that when it's given new set of training data.

Response variable can either be categorical or nominal. If classes are categorical, we identify problem as classification problem, if classes are nominal it belongs to regression. Classes can vary, depending on a number it is either binary or multi-label classification.

Unsupervised learning, we are only given inputs and goal is to find "interesting patterns" in data. Most common algorithms are clustering and outlier detection. In clustering algorithm data is grouped based on common characteristics. In outlier detection, main idea is to find such data that is so different from the whole data that it vastly changes whole scenery of data.

Reinforcement learning tries to find optimal behaviour or path to take into certain situation. Learning process is derived by constantly giving occasional reward or punishment signals [5].

2.2 Natural Language Processing

Natural language processing encompasses in various domains. Starting from machine transaction, converting text to speech, ending with OCR and sentiment analysis. Importance of NLP models in contemporary machine learning process is one the major fields.

Early steps of natural language processing started in the beginning of 1940s. The major problem that scientists were trying to tackle after the second world war was how to perform text translation in a machine way. One of the pioneers in the field was Noam Chomsky, his research showed first anomalies when grammatically correct sentences with no actual semantic value was equally irrelevant as grammatically incorrect ones.

Turning point in natural language processing was in late 1990s, when accessibility to internet and computers gave researchers chance to work much more data and direct their focus on information extraction and generation. Nowadays, NLP covers almost every area where human beings try to use text and speech in digital space, like text classification, mining, sentiment analysis, speech generation and classification.

2.3 Deep Learning

Deep learning is sub part of machine learning and is built up on artificial neural networks. One of the remarkable advancements in machine learning is neural networks that uses deep neural networking architecture. Neural networks like in human brains we have many connected neurons that transmit signals. Typically, neurons are organized into networks with different layers. An input layer usually receives the data input (e.g.: product images) and an output layer produces the ultimate result (e.g.: categorization of products) [6].

Traditional machine learning algorithms typically operate by defining a set of rules or features in the environment, these are hand engineered, human will look at data and try

to extract some hand engineered features from the data. In deep learning the key point is that these features are going to be learned directly from the data itself. E.g.: given a data to detect faces we can train a deep learning model to take as input a face and start to detect very low-level features and by building up those features we build eyes, noses and based on it we can build up larger features for face itself.

Today we are getting large data and by living in a world of big data we have a capability to use neural networks which are extremely parallelizable

2.3.1 Neural Networks

Building a block of neural network is started by creating single neuron, perceptron. We define set of inputs (x_1, x_2, …, x_m) this numbers are multiplied by their corresponding weights (w_1, w_2, …w_m) and then added together (see Figure 1). We take this single number that comes out pass it through a nonlinear activation function. In activation function we also add bias w_0, and we get the prediction $\hat{y}$. Formula: $\hat{y} = g(w_0 + X^T W)$. g sigmoid function maps output into [0, 1]. The simplest activation function which we will also use later in our training model is ReLu.

In a standard neural network architecture, we have fully connected layers, which means each input is taken from previous layer and is mapped to output.

Operation we apply to output neurons is SoftMax. The purpose of SoftMax is to turn the raw numbers that come out of a classification network into class probabilities [7].

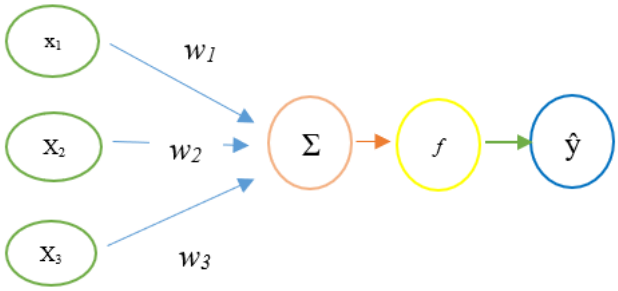

Figure 1. The perceptron: forward propagation

2.3.2 Convolutional Neural Network

One of the algorithms of deep learning is Convolutional Neural Network, the algorithm is commonly used in image processing. In the scope of this paper, it will be used for text classification. Before diving into the implementation, there are some algorithm specific components that needs to be explained.

One of the main components of CNN is kernels, sometimes called filters. Filters are square matrices that have dimensions $n_K \times n_K$, where n_K is an integer and is usually a small number, using kernels comes from classical image processing techniques [7].

Important aspect of CNNs is convolution. In the context of neural networks, convolution is done between tensors. The operation gets two tensors as input and produces a tensor as output [7].

Stride It's simply the number of rows or columns you move your region when selecting the elements.

Pooling layers are another kind of layer commonly used in convolutional neural networks. They simply down sample each of the feature maps created by a convolution operation.

To avoid the output shrinking as a result of the convolution operation, there is a method used throughout convolutional neural networks, "pad" the input with zeros around the edges, enough so that the output remains the same size as the input.

2.3.3 Convolutional Neural Network and Natural Language Processing

Techniques used for language processing varies from transferring raw waveform into spectral representation that can be used as an image, to single character level encoding. There are different ways of applying convolutions to text. Either applying it using world level representation or character level representation.

Word embeddings use text representation in a way that each single word is represented with a fixed dimension in a feature space. Words that are alike to each other will end up in the similar vector space. Example of such model is word2vec, where each word embeddings are concatenated to each other and then convolutions are applied across the

feature dimension and the variations of kernel size are along the time dimension. (See Figure 2).

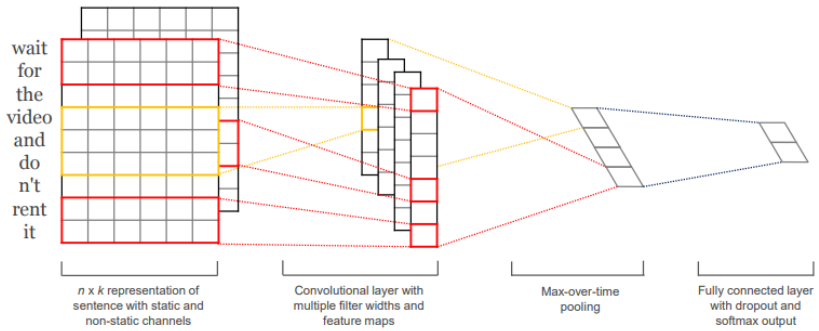

Figure 2. Model architecture with two channels for an example sentence.[22]

Character level model uses one hot encoding method meaning, already predefined alphabet, Unicode or ascii list is used as a character dictionary. Each letter is represented with each character and character is represented as a vector, with zeros and changed into ones when there is intersection on corresponding values.

	D	A	D
A	0	1	0
B	0	0	0
C	0	0	0
D	1	0	1
.	.	.	.
.	.	.	.
.	.	.	.
Z	0	0	0

Figure 3. One hot encoding

3 Research Background

Yearly analysis conducted by IWF shows growing risks for children facing online, especially girls aged 11-13 being targeted by sex offenders (See Figure 4). Most fearful is that more and more girls will become victims of the pernicious and manipulative forms of abuse. As Susie Hargreaves IWF CEO mentions “with more people spending more time online sex offenders find new ways to contact and manipulate children. time became pivotal and lockdown made it much worse” [9].

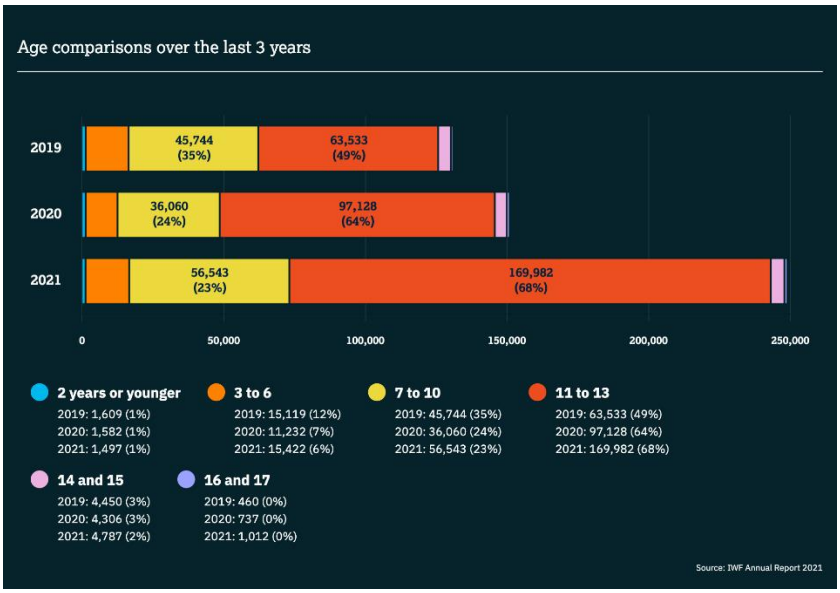

Figure 4. Age comparison of sexually abused victims. [9]

6 in 10 actioned reports specifically show the sexual abuse of an 11–13-year-old girl who has been groomed, coerced or encouraged into sexual activities via a webcam. 361 062 reports were assessed by IWF in 2020 out of 360 834 were reports of webpages and 228 were reports of newsgroups [9].

Here comes the question why the adolescents are most vulnerable group. One of the reasons might be that adolescents may be especially drawn to online relationships because of their intense interest in forming relationships, and because of the expansiveness of cyberspace frees them from some of the constraints of adolescence by giving them easy access to a world beyond that of their families, schools and communities [10].

3.1 Grooming children

Even though internet became publicly available late nineties sexual abuse of children has been already in place for centuries. First step, that is taken by abusers is grooming. The definition of sexual grooming is defined as followed:

"A process by which a person prepares a child, significant adults and the environment for the abuse of this child. Specific goals include gaining access to the child, gaining the child's compliance and maintaining the child's secrecy to avoid disclosure. This process serves to strengthen the offender's abusive pattern, as it may be used as a means of justifying or denying their actions" [11].

Tactics used by offenders are different, mostly predators adapt children's behaviour and groom based on aetiology. Child sex offenders have a special ability in recognizing vulnerable children [12]. Alternatively, offenders may target children or young people who have absent parents, or who were sexually abused as children because offender considers them easier to re-victimize [11].

3.2 Predators Linguistic approach

Linguistic approach tends to be one of the major aspects while analysing the online chats. It's a subject of research for a long time to detect common patterns predators are using while grooming children. When the purpose of self-representation is purely sexual, as in the soliciting of children, then the style of writing, use of punctuation, speed of response and use of emoticons will each influence the processing of first impressions [13].

In research of content analysis of the language used by offenders, which was conducted by pretending being child in online chat rooms, authors emphasize eight recurrent themes that reflected in conversations. The analysis begins with a representation of the raw content of the sexual offenders' discourse, distinguishing between 'implicit and 'explicit' comments, this relates to second category which looks at the way in which sexual offenders appear to 'solicit' children for sex on-line [14]. The methods of solicitation used by sexual offenders' 'initiation' and 'transference' tie to the objective whilst on-line. This theme is referred to as 'fixated discourse, any deviation from the principal focus embraces the themes of 'conscience' and 'colloquial language'. Following this, the analysis

examines how sexual offenders appear to ‘minimise possible risks of detection; and how they ‘acknowledge illegal/immoral behaviour’ [14].

Later in the paper these eight language themes will be used to evaluate linguistic patterns in research data. Since these recurrent patterns can be used to connect conversations to potential abusers, it will be interesting how these themes happen in different set of data.

3.3 Online predator identification with classification

One of the first approaches to identify online predators by applying classifiers was accomplished in early 2007. The aim of the paper was to automatically distinguish between pseudo-victim and the predator, research showed very high accuracy and trained series of SVM and distance-weighted k-Nearest Neighbour (k-NN) classifiers were used for training purposes [15]. Later first deterministic approach to label and analyse chat transcripts was done by April Kontostathis in 2009. The system used a rule-based approach in conjunction with decision trees and k-NN classifier as an instance-based learning method [16].

Classification algorithms that have been used various from k-Nearest Neighbour (Kanger, 2012), Entropy-based Classification (Eriksson and karlgren, 2012), Support Vector Machine (Morris, 2013) and Neural Networks (Villatoro-Tello et al., 2012).

Later, sequences of classifiers were used for predator detection in chat conversations based. Chain classifiers were used for three stages employed by predators to approach the victim [17]. After simply analysing chat logs with classifiers research have been done which use high-level features and evaluating their applicability in detection process. Listing the features, including sentiments as well as other content-based features. Stated idea is that emotional markers such as feelings of inferiority, isolation, loneliness, low self-esteem and emotional immaturity is common for sexual predators who approach minors [18]. These markers were used as features with SVM classifier and method showed 94% accuracy.

4 Methodology

4.1 The data

The corpus for this study was taken from PAN competition which was conducted in 2012. Aim of this competition was to at providing research with a common framework to test methods for identifying such misbehaviours or cybercriminal activities. Entire data is in English language and corpus contains 4 671 049 lines, all individual conversations are (n=66 927) out of 64911 is non-predatory and 2016 - predatory.

The data itself that was used in a competition is collected from two major source: http://www. perverted-justice.com/ (PJ) website. This is a website where logs of online conversations between convicted sexual predators and volunteers posing as underage teenagers are published, also, it includes IRC logs thousands of conversations, that were already made available on the website of the IRC channel managers, namely http://www.irclog.org/ and http://krijnhoetmer.nl/irc-logs/ [19].

4.2 Framework

The paper adopts Corpus-Assisted Discourse Studies methodology in the first part, later deep learning, specifically CNN will be used to make detect predatory chats.

Firstly, analytical part will be conducted to generate distributional information about the corpus. The study about content analysis of the language used by offenders will be used as a base and data will be evaluated based on eight themes: ‘implicit/explicit content’, ‘on-line solicitation’, ‘fixated discourse’, ‘use of colloquialisms’, ‘conscience’, ‘acknowledgement of illegal/immoral behaviour’, ‘minimizing the risk of detection’ and ‘preparing to meet offline’. Also, frequently – ranked wordlists will be identified which statistically tends to occur. The process also includes extensive reading of corpus and software tools that compute statistical information. Specifically, in this part (n=200) predatory conversations will be used. These conversations are selected based on following criteria: conversation should be long enough to detect at least one theme.

In the second part of research convolutional neural network algorithm will be used to detect predatory conversations. Worth mentioning, the algorithm has already been used

in 2012 during PAN competition, but this paper will take a reference from Character-level Convolutional Networks for Text Classification by Xiang Zhang, Junbo Zhao, Yann LeCun, Paper was published in 2015 and compares large data corpuses against traditional models such as bag of words, n-grams and their TFIDF variants, and deep learning models such as word-based ConvNets and recurrent neural network.

4.3 Procedure

4.3.1 Data extraction and formatting

The chat logs that were used in PAN competition is in XML format containing all predatory and non-predatory conversations together (see Figure 5), separately data includes text file of Ids of predator authors. Which will be later used to divided and label data. In order to keep data as much close to original, only blank lines were excluded from corpus, other than that no modifications have been done.

```
<conversation id="e621da5de598c9321a1d505ea95e6a2d">
  <message line="1">
    <author>97964e7a9e8eb9cf78f2e4d7b2ff34c7</author>
    <time>03:20</time>
    <text>Hola.</text>
  </message>
  <message line="2">
    <author>0158d0d6781fc4d493f243d4caa49747</author>
    <time>03:20</time>
    <text>hi.</text>
  </message>
  <message line="3">
    <author>0158d0d6781fc4d493f243d4caa49747</author>
    <time>03:20</time>
    <text>whats up?</text>
  </message>
  <message line="4">
    <author>97964e7a9e8eb9cf78f2e4d7b2ff34c7</author>
    <time>03:20</time>
    <text>not a ton.</text>
  </message>
```

Figure 5. Conversation Sample

```
00851429b21722a4d62f63a328c601ca
00aac10b39157377c79b7700b7b832bf
02800e11fdb1b43595303709f2b38f8c
03957f443c7790f9642db14bbc59df11
04bfa707d3313179ef48177d7270938e
04d42f7bb1eb41605dea74a8711f9fd0
0526eb9cfcee11c0036f3fa6d11158d5
053364a8ce3df76dadd5fe75fb056f72
0599cd3f7fc15849844468b0702ff593
0b6b05c740a1bf50ca7f9a461598a3b9
0cac6dfd241c5efe4ca07417575e582f
10e49cbbe257b19162a677113236cdc2
12140a644bf57166fe014116d1761ac0
1290ea419856093edf96b1263cb1ca1e
```

Figure 6. Predatory Ids

The research doesn't take analysis of message time, message line. Therefore, only text lines are scrapped from XML file and based on ids divided into predatory and non-predatory text files.

Firstly, predatory Ids are extracted from text file and list is created (see Figure 7).

```
import os
f = open("predatoryIds.txt", "r")
list = f.read().split()
```

Figure 7. create predatory id list

Figure 8 shows data extraction and division into predatory and non-predatory chats. Complete code is provided in Appendix 1

```
for conversation in root:
    if conversation.tag == 'conversation':
        print(conversation.attrib['id'])
        converdsationId = conversation.attrib['id']
        totalConvCount += 1
        for author in conversation.iter('author'):
            convid.append(author.text)
        check = any(item in convid for item in list)
        if check:
            predatorylaceholder = open(
                'cleanedData/predatory/'+conversation.get('id')+'.txt', 'w')
            for attr in conversation.iter('text'):
                print('*** found predatory****')
                predatorylaceholder.write(attr.text)
            countPredatory += 1
        else:
            nonpredatorylaceholder = open(
                'cleanedData/nonpredatory/'+conversation.get('id')+'.txt', 'w')
            for attr in conversation.iter('text'):
                print('not predatory')
                nonpredatorylaceholder.write(attr.text)
            countNonPredatory += 1
        convid.clear()
```

Figure 8. Data extraction and Division

In order to train data, proper labelling was necessary. Data was created as a data frame where each conversation is labelled as either predator or non-predator (see Figure 9).

```
d = {'X': [], 'Y': []}
for filename in os.listdir(directory):
    print(filename)
    alltext = open(directory+'/'+filename)
    print("**************************")
    allitext = alltext.read()
    d['X'].append(allitext)
    d['Y'].append(PREDATOR)
for filename in os.listdir(directorytwo):
    print(filename)
    alltext = open(directorytwo+'/'+filename)
    print("**************************")
    allitext = alltext.read()
    d['X'].append(allitext)
    d['Y'].append(NONPREDATOR)

with open('dataorganaized.json', 'w') as jsonfile:
    json.dump(d, jsonfile)
```

Figure 9. Labeling data

4.3.2 Tools

Qualitative language content analysis was conducted in software NVivo 12. Program was mainly used to code the conversations with themes and to perform cluster analysis. Also, program was used to get frequently managed word lists and colocations. Tool is particularly useful for facilitating advanced analysis and for finding evidence-based insights faster.

Convolutional neural network algorithm was implemented in Google Colab. Environment provides free resources for machine learning models and has some modules already pre-installed.

Google Colab machine specifications: NVIDIA -SMI 460.32.03, Driver version: 460.32.03 VM: Tesla K80. Cuda compilation tools, release 11.1

In order to prepare environment fully. MXNet framework was installed for deep neural networks training. For compatibility with cuda version additional packages were installed.

```
!pip3 install torch==1.9.0+cu111
torchvision==0.10.0+cu111 torchaudio==0.9.0 -f
https://download.pytorch.org/whl/torch_stable.html
!pip install mxnet-cu110
```

Figure 10. package installation

5 Results

5.1 Linguistic Content Analysis

Looking into themes sparse results and their interconnection (see Figure 11) we grab interesting relation between themes. How The words used in each theme interrelates. As code analysis show preparing to meet offline and use of colloquialisms are in the same subgroup that can be indicator of texting positive responses to appear more appealing to target is a ground preparation for gaining trust and increasing sympathy for potential meeting.

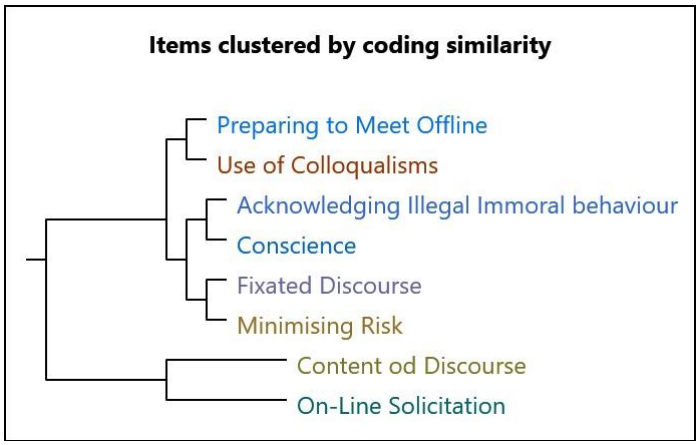

Figure 11. Items by coding similarity

Acknowledging illegal/immoral behaviour and consciousness being related are only similar in case of empathic conscience where, offender acknowledges that his response might cause negative or distressing impact on victim.

During experimentation also word frequency was evaluated, stop words (e.g.: the, a, is, are) were excluded from analysis. Minimum length for word was defined as 3 and frequency analysis included stemmed words (e.g.: talk, talking, talked). In a separate run including synonyms were tried, but results weren't useful. Words like: break, breaks, clothes, clothing, wear, tired was considered as a synonym to wear.

Word frequency showed highest percentage of for word lol 3,41%, wants 1,58 % meet 0.34% coming 0.59%. For assumptions purposes only words: 'meet', 'want', 'alone',

‘home’ were taken. Even thought, some words had higher portion, it doesn’t carry much context for analysis, so they were excluded.

Figures 12, 13, 14 show each word’s contextual representation.

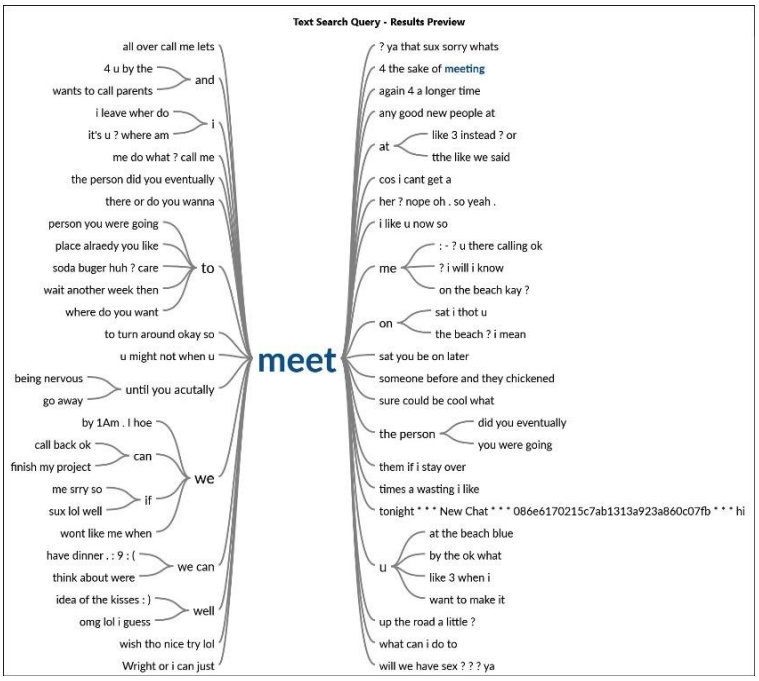

Figure 12.

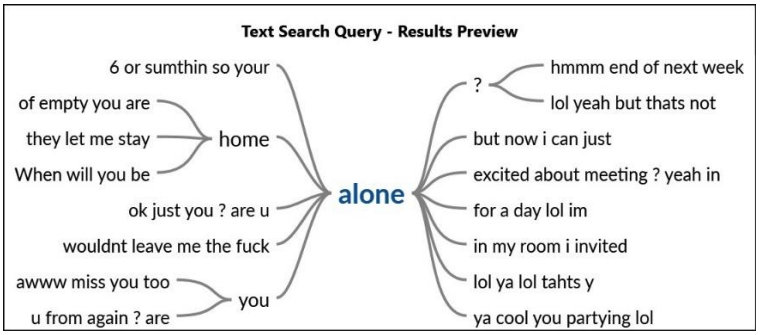

Figure 13

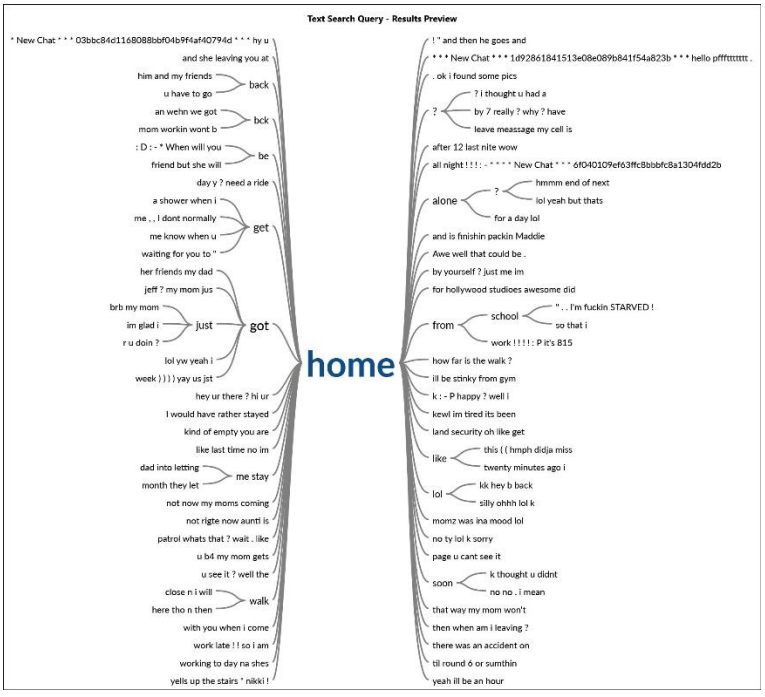

Figure 14

5.2 Character Level Convolution

Model accepts a sequence of encoded characters as an input. The encoding is done by prescribing an alphabet of size m for the input language, and then quantize each character using 1-of-m encoding (or “one-hot” encoding), then, the sequence of characters is transformed to a sequence of such m sized vectors with fixed length l_0, and any character exceeding length l_0 is ignored, and any characters that are not in the alphabet including blank characters are quantized as all-zero vectors [20].

Network consists of 6 convolutional and 3 fully connected layer (see Figure 13). According to the paper the input has number of features equal to 70 and feature length is 1014, 2 drop-out layers in between 3 fully connected layers with 0.5 probability rate [20].

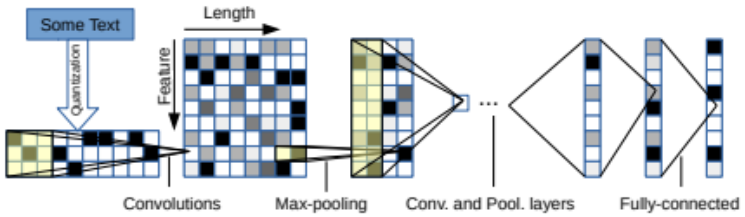

Figure 15. Network Creation [20]

Figure 14 shows detailed information about convolutional layers.

Layer	Large Feature	Small Feature	Kernel	Pool
1	1024	256	7	3
2	1024	256	7	3
3	1024	256	3	N/A
4	1024	256	3	N/A
5	1024	256	3	N/A
6	1024	256	3	3

Figure 16. Convolutional Layers

Model training purposes on a first iteration of experiment network is build based on the parameters provided in the paper. Figure 17 entails details of training implementation.

```
for e in range(start_epoch, number_epochs):
    for i, (review, label) in enumerate(train_dataloader):
        review = review.as_in_context(ctx)
        label = label.as_in_context(ctx)
        with autograd.record():
            output = net(review)
            loss = softmax_cross_entropy(output, label)
        loss.backward()
        trainer.step(review.shape[0]
        curr_loss = nd.mean(loss)
        moving_loss = (curr_loss if (i == 0)
                                    else (1 -
            smoothing_constant) * moving_loss +
            (smoothing_constant) * curr_loss)

        if (i%100 == 0):
            print('Batch {}: Instant loss {:.4f}, Moving
loss {:.4f}'.format(i,curr_loss.asscalar(),
moving_loss.asscalar()))
```

Figure 17. Model Training [21]

After training model based on parameters provided, precision of model is nearly 70%. Important note from the paper is that it uses maximum 1014 characters per document, which in our case is too small number. Most of the texts are longer than 1014 characters.

Following experiments examines the different parameters, with increased number of character size. Evaluation parameters for these experiments are: Accuracy, Recall, Precision and F_1 score.

Accuracy is the total number of correctly predicted samples. Which is calculated by following formula: $Accuracy = (TP+TN)/(TP+TN+FP+FN)$.

Recall is the proportion of correctly predicted data as it was classified. $Recall = TP/(TP+FN)$

Precision is proportion of total positive predictions that, where actually positives. $Precision = TP/(TP+FP)$

F_1 score is a harmonic mean of recall and precision. Since, training data is highly skewed harmonic mean is used for evaluations. $F1 = 2*precision*recall/(precision+recall)$.

- TP (True Positive) - number of predator samples as predicted predators
- FP (False Positive)- number of nonpredator samples that are predicted as predators
- TN (True Negative) – number of nonpredator samples that are predicted as nonpredator
- FG (False Negative) – number of predator samples that are predicted as nonpredator

Table 1 shows results for different Kernel size for each convolution layer and numbers of filters applied. Best result comes with higher number of filters and kernel size starting from 7.

Number of filters	Kernel size -layer	Accuracy	Precision	Recall	F_1-score
1024	7,7,3,3,3,3	0,51	0,52	0,42	0,46
256	5,5,3,3,3,3	0,49	0,21	0,10	0,14
256	9,9,3,3,3,3	0,50	0,51	0,37	0,43

Table 1. Model experiments for different parameters

6 Summary

The problem, online sexual abuse, is one of the important problems in today's cyberspace. Unfortunately, there is a paucity of empirical data on how juveniles are groomed for potential sexual abuse via internet. Scarcity of data is due to the regulatory and sensitivity issues. The paper main goal was to show how machine learning tools can be used to support digital forensics for their investigations. In the following paper we tried to show one of the methods of Machine learning and show the results.

Through training data with ML models, linguistic content analysis was done, themes and recurrent patterns was identified. Analysis was also done for most common words used in chats and contextual mapping for the words. Based on results limitations of models can be assessed.

In future works, much more datasets labelled and provided by forensics experts will help to generalize results and have much more precision when conducting analysis and research. This, in return, will provide crucial findings that will benefit to children protection.

References

[1] M. Motivans, T. Kychelhan “Federal prosecution of child sex exploitation offenders”, 2006

[2] Sextortion: Online threat to kids and teens [online]. Available:

https://www.fbi.gov/scams-and-safety/common-scams-and-crimes/sextortion

[3] Increased amount of child sexual abuse material detected in Europe [online]. Available: https://ec.europa.eu/home-affairs/news/increased-amount-child-sexual-abuse-material-detected-europe-2020-04-28_en

[4] McKean, Erin. The New Oxford American Dictionary. New York, N.Y: Oxford University Press, 2005. Print.

[5] T. M. Mitchel, “Machine Learning”, 1997, pp. 2.

[6] P. Zschech, K. Heinrich, Ch. Janeish “Machine learning and deep learning”, 2021

[online] Available: https://link.springer.com/article/10.1007/s12525-021-00475-2

[7] A. Glassner “Deep learning : a visual approach”, 2021.

[8] U. Michelucci “Advanced Applied Deep Learning: Convolutional Neural Networks and Object Detection”, 2019, pp 78-79.

[9] IWF Annual Report [online]

Available: https://annualreport2021.iwf.org.uk/trends/

[10] Wolak, Mitchell, Finkelhor. “Escaping or connecting? Characteristics of youth who form close online relationships”, 2003.

[11] S. Craven, S. Brown, E. Gilchrist “Sexual grooming of children: Review of literature and theoretical considerations”, 2006.

[online] Available: nationalcac.org/wp-content/uploads/2019/05/Sexual-grooming-of-children-Review-of-literature-and-theoretical-considerations-Craven-2006.pdfn .

[12] J. R. Conte, S. Wolf, T. Smith, “What sexual offenders tell us about prevention strategies. *Child abuse and neglect*”, 1989

[13] F. Mantovani. “Networked seduction: a test-bed for the study of strategic communication on the Internet. CyberPsychology & Behaviour, 4, pp 147-154

[14] V. Egan, J. Hoskinson, D. Shewan. “Perverted Justice: A Content Analysis of the language used by offenders detected attempting to solicit children for sex”, 2011.

[15] N. Pendar, "Toward Spotting the Pedophile Telling victim from predator in text chats," International Conference on Semantic Computing (ICSC 2007), 2007, pp. 235-241, doi: 10.1109/ICSC.2007.32. [online]

Available: https://ieeexplore.ieee.org/abstract/document/4338354

[16] A. Kontonstathis, “ChatCoder: Toward the Tracking and Categorization of Internet Predators”, 2009.

[17] H. Escalante, A. Juarez, “Sexual Predator Detection in Chats with chained classifiers,2013. [onlne] Available: https://aclanthology.org/W13-1607.pdf

[18] D. Bogdanova, “Exploring high-level features for detecting cyberpedophilia”, 2014 [online]

Available: https://www.sciencedirect.com/science/article/pii/S088523081300034X

[19] G. Inches, F. Crestani “Overview of the international sexual predator identification competition at PAN-2012”

[20] X. Zhang, J. Zhao, Y.LeCun, “Character-level Convulationa Networks for Text Classification”, 2015.

[21]
https://github.com/ThomasDelteil/TextClassificationCNNs_MXNet/blob/master/Crepe-Gluon.ipynb

[22] Y.Kim “Convolutional Neural Networs for Sentence Classification”, 2014

Appendix 1 – Data Division and Extraction Code

```
import xml.etree.cElementTree as ET
from filedetect import list

tree = ET.ElementTree(file='test5.xml')
root = tree.getroot()
countPredatory = 0
countNonPredatory = 0
totalConvCount = 0
convid = []

for conversation in root:
    if conversation.tag == 'conversation':
        print(conversation.attrib['id'])
        converdsationId = conversation.attrib['id']
        totalConvCount += 1
        for author in conversation.iter('author'):
            convid.append(author.text)
        check = any(item in convid for item in list)
        if check:
            predatorylaceholder = open(
                 'cleanedData/predatory/'+conversation.get('id')+'.txt', 'w')
            for attr in conversation.iter('text'):
                print('****reahced the place found predatory****')
                predatorylaceholder.write(attr.text)
            countPredatory += 1
        else:
            nonpredatorylaceholder = open(
              'cleanedData/nonpredatory/'+conversation.get('id')+'.txt', 'w')
            for attr in conversation.iter('text'):
                print('not predatory')
                nonpredatorylaceholder.write(attr.text)
            countNonPredatory += 1
        convid.clear()
```

Appendix 2 – Create Data Frames Code

```
import os
import json
from collections import defaultdict
PREDATOR = "predator"
directory = 'cleanedData/predatory'
NONPREDATOR = 'nonpredator'
directorytwo = 'cleanedData/nonpredatory'
# items=()
d = {'X': [], 'Y': []}
# print(type(items))
for filename in os.listdir(directory):
    print(filename)
    alltext = open(directory+'/'+filename)
    # print(alltext)
    print("**************************")
    allitext = alltext.read()
    # print(allitext)
    d['X'].append(allitext)
    d['Y'].append(PREDATOR)
for filename in os.listdir(directorytwo):
    print(filename)
    alltext = open(directorytwo+'/'+filename)
    # print(alltext)
    print("**************************")
    allitext = alltext.read()
    # print(allitext)
    d['X'].append(allitext)
    d['Y'].append(NONPREDATOR)

with open('dataorganaized.json', 'w') as jsonfile:
    json.dump(d, jsonfile)

print(d['Y'])
print(d['X'])
```

Appendix 3 – Non-exclusive licence for reproduction and publication of a graduation thesis[1]

I Nani Khantadze

1. Grant Tallinn University of Technology free licence (non-exclusive licence) for my thesis “Detecting Online Sexual Predatory Conversations in Minors”, supervised by Alejandro Guerra Manzanares
   1.1. to be reproduced for the purposes of preservation and electronic publication of the graduation thesis, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright;
   1.2. to be published via the web of Tallinn University of Technology, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright.
2. I am aware that the author also retains the rights specified in clause 1 of the non-exclusive licence.
3. I confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights, the rights arising from the Personal Data Protection Act or rights arising from other legislation.

16.05.2022

1 The non-exclusive licence is not valid during the validity of access restriction indicated in the student's application for restriction on access to the graduation thesis that has been signed by the school's dean, except in case of the university's right to reproduce the thesis for preservation purposes only. If a graduation thesis is based on the joint creative activity of two or more persons and the co-author(s) has/have not granted, by the set deadline, the student defending his/her graduation thesis consent to reproduce and publish the graduation thesis in compliance with clauses 1.1 and 1.2 of the non-exclusive licence, the non-exclusive license shall not be valid for the period.