1st International Conference on Trends in Advanced Research

March 4-7, 2023 : Konya, Turkey

All Sciences Proceedings

http://as-proceeding.com/

Machine Learning and Artificial Intelligence-based Child Abusing Tracking System for the Detection of Online Sexual Predators

Zeeshan Ahmad[1], Umut Özkaya

[1]Department of Computer Science, SEST, Jamia Hamdard, New Delhi, India
[1]Department of Electrical and Electronics, Konya Technical University, Konya, Turkey

zahmad090@gmail.com Email of the corresponding author

***Abstract* –** With the upward thrust of cybercrime targeting kids, chat logs may be used to hit upon and mark harmful conduct for law enforcement. Children can be helped by this in a great manner from cybercrime. Previously digital forensic investigations were usually done by hand; this traditional approach of relying on the assessment was not reliable. The solution proposed in this paper uses the Digital Forensic Model using machine learning and artificial intelligence-based on various Child Abuse Tracking System supported by Microsoft technology and other companies to facilitate automatic detection of harmful conversations on the chat log. Therefore, the contribution of this paper is to show how the activities in the digital investigation process can be organized to obtain useful results using PhotoDNA technology (integrated CETS) helping law enforcement, fight child pornography when investigating online attackers. In addition, this paper included the study of Artemis, an automated system that scans chats to detect online sexual predators. So far no one has done any study on both of these tools. We have proposed architecture for the detection or efficiently capturing of the predators by enabling advanced technologies like machine learning, artificial intelligence, geographical information systems, and data mining. This architecture is based on the category of child exploitation and the scenario of the integrated model based on the material collected. The collaboration of all these aspects in an updated and efficient manner can come up with an effective result and can help law enforcement to take necessary actions.

Keywords – Machine Learning, Artificial Intelligence, Geo-Information System, Digital Forensic, Cyber Security, Online Sexual Predatory Chats.

I. INTRODUCTION

As per the recent reports, exploitation of children has become a major problem across the globe, we need a smart, efficient tracking system that helps us to track or monitor the child's data efficiently. In this model, we have created architecture with the help of this architecture we describe the overall working of the system. This Research paper aims at determining the role of ML and AI in tracking chat logs, for this we have a synthesized and updated version of the Child Exploitation Monitoring system which works similar to the Microsoft monitoring system but it has many updated advanced features in that system such as this system has advanced monitoring using machine learning and artificial intelligence, we can keep an eye on children's activities through GIS as a decision support tool and track the data or secure the data using blockchain monitoring. Blockchain helps us to secure the data so that no mediator or dealer steals the data of the children so that our data exploitation chances of children become less and

by using Google maps or geographical information systems we can track the children. It examines how big data analysis can be effective in tracking and preventing the harmful behavior of predators. In this, we used advanced tools to gather information, so they can identify, prevent, and punish those who commit crimes against children. CETS is designed to simplify investigations and integrate other CETS activities so that law enforcement agencies can improve their cooperation and advance their cases.

A variety of strategies have been developed to design smart systems to solve various digital forensics problems in the field of information security. Machine learning (without human interference) can collect, analyze, and process data. AI and ML, with these two tools, researchers can create their processes in digital forensics. This can help the inquisitor to put the task on a self-directed mode MLF with help of AI as it can produce a great deal of data analysis that can help to determine and detect if any activity is safe or unsafe, has criminal and dangerous traits, or normal, weather it has moral issues or it is apt for them. Online bullying training is a strategy or set of tactics designed to make a child trustworthy and forced to be sexually abused. If there is no supervising body with the child, this situation can be dangerous and can lead the child to sexual abuse or even to trafficking. This is a method that captures text-based discussions for redress, allowing individual companies to monitor potential corrections and report them to law enforcement. It will help to prevent children from sexual abuse and enable organizations to report and correct themselves in the courts where it has been done. Online Child abuse is a heinous crime that requires a firm public reaction.

II. Problem Statement

It is no secret that, although the Internet has become the largest and most effective connector for people all over the world, it has also become a threat to society as it has a lot of negative impacts also. One of the most significant negative effects is the dark relationship between the child sexual abuse industry and the internet. Before the advent of the internet, it was very difficult and very dangerous for sexual predators to find potential victims and sell them to clients.

Now, with the advent of technology, finding and selling people is just as easy as a few clicks. And while it is not a guarantee that all the kids online will encounter an attacker who wants to kidnap or physically control and exploit them, it does happen. To find potential victims, abusers now can anonymously identify vulnerable young people by posting their profiles on social media, contacting those using inappropriate or false profiles, and initiating a self-improvement process. In addition, abusers can build trust in what they intended to happen which is not possible face-to-face; they try to win the trust of the child by praising, helping, and doing other things like that to make the child feel special.

On the other side of the coin, sexual predators may force their victims to take action by gaining power over them. The rapist acquires this power by monitoring them or monitoring their accounts, sending threat messages, spreading rumors about their victims, or by blackmailing them for any reason. When it comes to selling victims, it is as easy as uploading a craigslist ad. Once they have contacted the child, abusers will find “dating” and “personal” sites that offer more than what they suggest, and then upload an ad with a name, a brief description, a phone number, and photos exposing their victims [1]. Thankfully, the world has started to recognize and begin to respond to such activities.

III. Background

A lot of studies have been conducted regarding the exploitation of children. According to the studies, the major amount of social networking sites is proven to be the highly used tool for the abuser to abuse the child. Bad guys or we call them predators, use social networking sites, and start blackmailing the children which cause many hazardous impacts on children and society as well. Online sites have fewer foundations and it is very easy to do friendships and built trust with children through online networking sites, so we can consider this the main reason behind child sexual abuse [2]. There are lots of sexual predators who are always online and they start blackmailing children using the stuff which was shared by the victims with predators based on trust.

Given below is a record of randomly selected online material on the ICSE database, INTERPOL, and ECPAT International; it was published in a joint report in February 2018. The research notified many shocking trends:

- 84% of photos had explicit sexual content.

- More than 60% of unknown victims had not yet been identified, including infants and toddlers.
- 65% of unknown victims were girls.
- Extremely violent scenes may expose boys.
- 92% of the offenders were men.

India had reported more than 24 lakh incidents of child sexual abuse or harassment cases via online mode in a period of three-year. As per Interpol data from 2017 to 2020, an average of 80 percent of the victims are girls who are under the age of 14 years. As shown in fig-1, the figures suggested CBI launch a great crackdown on accused of online child sexual harassment based on CSAM in India through SNSs under their responsibility to handle such incidents. According to Interpol data CSAM content and consumers are growing exponentially, evidences say that 1.16 lakh inquiries are raised on child pornography in a single Internet search engine. The central probe agency plans to take up the matter with the help of SMPs through the apt legal provisions to assess at their end. It depicted those 2.4 million incidents of child sexual abuse were online and 80 percent of the victims were young girls under the age of 13 years who are proven to be worthy victims.

Source: EVAC

Fig-1: Statistics of worldwide cases of online sexual exploitation and abuse

IV. RELATED WORK

Previously authors [3] described the job evaluation framework which provided concrete evidence that the CNN-based model is strong against criminals. Efforts to invest in the development of metadata-based vision systems can increase Child Sexual Abuse Material(CSAM) detection rates and can help thousands of victims. The framework provides guidelines on how to test the CSAM acquisition model against intelligent enemies and guidelines on how to test CSAM recovery models with open data. With such barriers, the development of CSAM machine learning structures based on record metadata opens many chances. Later on, the authors [4] focused on the task of finding someone who takes drugs early (eSPD). An important danger, children face today is self-correction online, in which the so-called abuser establishes an online contact with a young child to sexually abuse them to resolve this new data called PANC to test reality.

Earlier authors [5] centered to demonstrate how the movements within the digital investigation technique can be organized to reap useful ML results while investigating predators online. The Digital Forensic system version supported by ML techniques helps the automatic detection of dangerous conversations within the chat log [6]. Generally, online sex acts were considered “normal” among teenagers, although they were considered “disgusting”. There are several types of child exploitation that define the severity of abuse based on online activity like sexually explicit material, child pornography, and live video games. Reference [7] discusses the problem of cyberbullying online by opening words and emotional behavior, which can help identify if a user is a good person or an attacker. To strengthen the acquisition model, the paper also focuses on dividing the two categories of users based on the emotions expressed in their conversations. Two data sets are created in terms information-based information and emotion-based information. Previously authors [8] have aimed to test the ability of a malicious character to reveal another user's personal information by identifying artifacts that may refer to sensitive user data. Since while using geolocation features applications are intended for users to provide personal information when conducting their search to meet someone, the same information may be used by cybercriminals or forensic analysts to gain access to personal data.

V. PROPOSED METHODOLOGY

This paper relies on synthesizing research findings of previous studies and tools conducted on the

role of ML and AI in assessing challenges. In our proposed solution we use three major aspects to extract the original information, these three important aspects are:

1-Geographical Information System

2-Data Mining

3-Machine learning and artificial intelligence: PhotoDNA, Artemis

We integrated these three aspects in the first part. As fig-2 shows, we categorize the data into four categories to evaluate which category the data belong to and what are the mechanism and aspects we can apply. As we get the information about the data, we can use these three aspects to track predators. A Geographic information system is used for tracking purposes or for finding the location. The second aspect is one of the major parts which are used for data extraction and only the important data related to this is being extracted, the third one is the most important tool in our proposed solution which is machine learning and artificial intelligence. In this part we give or train the machine in a supervised learning manner wherein artificial intelligence helps us to work smartly or effectively so the integration of all of these constitutes an advanced system that is very helpful for the tracking and reducing the exploitation of children and the main motive of this research is to protect the society from negative people and this system will help us to maintain peace and children can enjoy their life without fear of exploitation.

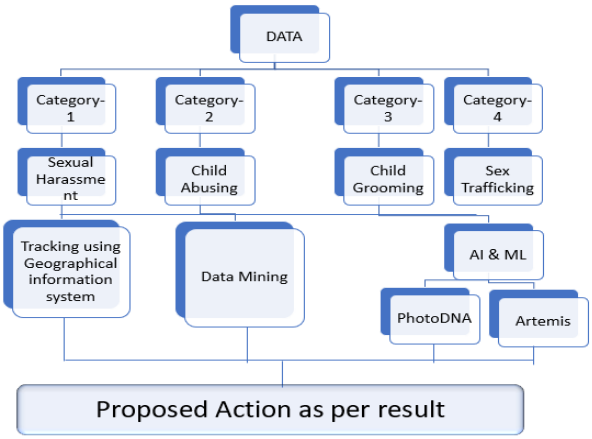

Fig-2: Architecture of Child Abusing Tracking System

Sexual Harassment: As you know that child abuse is increasing day by day and the attachment or involvement of children with social networking sites is growing, and cyberbullying is any unwanted online behavior. A child can be abused, feel scared, or humiliated by online sexual abuse or you can say harassment, it can be in various ways. It can be through pornographic videos, photos, obscene messages inappropriate posts or pages, etc.

Child Abusing: Online child-abusing means vulgar talk and use of inappropriate language which is not suitable and should not be done with the children. During the Covid pandemic, kids started using mobile phones more than ever sometimes parents provide them with mobiles for playing games to avoid getting bored and sometimes for study purposes in this way the chance of cyberbullying or abuse increased that occurs deliberately on the internet through SMPs, online gaming and SNSs. Spending more time with mobile phones using the internet proved to be a golden opportunity for the predators to abuse the child

***Child Grooming*:** For setting up relationships with the child, the internet helps the predators and provides them the opportunity to abuse sexually or blackmail the child using digital technology, sometimes they threaten or impress the child to the extent that the child came into their contact offline as well then, these predators get chance to abuse the child physically. Another misuse of social networking sites is to train children for sexual misconduct. Predators make a relationship for the benefit purpose or blackmailing the child; this type of work and activity is done in the child grooming.

***Sex Trafficking*:** When young people are engaged in commercial sex, even when there is no overpowering, no fraud or coercion, they are considered victims of human trafficking. Sex traffickers remain active on SNS to find their victims and they can find them also easily as people do not understand the consequences and risk of trusting such sites and fake people, sometimes the information provided to these sites leads to life threats also, the predators start kidnapping the children who are below 18 years of age and start sex trafficking with these children, these types of activities are very harmful to the children and for the society also.

GIS: Through GIS we can collect, store, and analyze data digitally and can get the records

anytime. For convenience, GIS may be used to display and interact with digital maps. They may be used to analyze and extract records from nearby records, using neighborhood energy to recognize what's taking place and wherein. GIS is a system that helps to find or track the children or predators' location after that we can get the children free from predators and it became a very efficient and good step for finding the data related to the children and when we get the information of predators, we send the location to the nearest law enforcement office and after that officer can catch the predators very easily and save the child from predators. We can use a decision support tool with this to get the nearest location and which helps us to make a decision.

As you know, the geographical information system is the fastest emerging technology that's why we used it here as it is useful for tracking purposes and to track the predators as well as the children, which is the major work in our system.

Data Mining: Data Mining is a tool for finding patterns, external discovery, or information from huge amounts of data. The method of extracting statistics to perceive patterns, developments, and beneficial statistics that could permit an entity to make a records-driven choice on large statistics sets is referred to as Data Mining. In other words, we can say that Data Mining is a procedure of investigating hidden patterns of data in exclusive views for classification. Data mining plays a very important role in categorization. With the help of this, we can categorize the data in such a format that we can easily gather the related information of the children as discussed above. Data plays a very vital role in this but in big data analysis we may collect lots of insufficient data also, so data mining is very important after that we can collect only the relevant data which is helpful to collect the information related to predators, the data related to predators is essential to capture or track the predators in a fast manner.

As we know that we can easily put the command and queries when the data is extracted and get the actual information related to the particular task where it is needed, in this way the piece of information which is not required now will be stored and can be used in future whenever it is required.

ML & AI: These are the tools defined as obtaining knowledge and skills along with this Artificial intelligence can utilize that data for future context. ML allows the system to learn new things from data, leading to the development of a simulation system to respond to a particular behavior in a situation. It involves creating self-study algorithms.

In the present scenario, MI and AL are key factors and play a pivotal role so we are using both of them for our study. With the help of machine learning, we can learn the machine and can track the predators 24*7 and artificial intelligence helps us to define the acquisition of knowledge and intelligence which is helpful to acquire skills in a very efficient and effective manner.

With the help of artificial intelligence, we can detect online predators; here are some aspects or activities which are done by the victim [9].

- Becomes secretive about online activities
- Becomes obsessive about being online
- Becomes indignant while he or she does not get the opportunity to be online
- Receives phone calls from people you do not know or makes calls to numbers that you do not recognize
- Receives gifts, mail, or packages from someone you do not know
- Withdraws from family and friends
- Changes screens or turns off computer when an adult enters their room
- Begins downloading pornography online

These are the major things that can be observed later the child starts to be abused by predators mostly in the beginning predators start blackmailing or misuse of children.

Tools which are supported by ML and AI are very important to track predators, so we used machine learning and artificial intelligence in our study, specifically PhotoDNA and Artemis tools. These tools are very helpful to track and reduce child exploitation.

A. PhotoDNA

Having worked around the world for many years, there is always a great need for research tools and resources. One such tool is PhotoDNA, which creates something like photo fingerprints that can be compared to other image signatures to

obtain copies. This includes CETS and which has some set of laws like POCSO which helps to track and impose laws, it has become easier to catch hold of such criminals or suspects so that victims should get justice without being troubled, acquire evidence and prosecute kids suspected of having pornography.

Use of PhotoDNA to fight against online child abuse:

Because people who share child pornography videos often embed this illegal content in a flick, cartoon, or innocent home movie, it may take up to a half-hour or some hours to find the content material and determine which video must be downloaded and reported for regulation of enforcement. Recently, IWF, an international watchdog, started using PhotoDNA, a tool developed by Microsoft in 2009 for standalone video identification of child pornography [10].

PhotoDNA also allows content material that helps to get rid of thousands and thousands of illegal pics from the net; assists in the conviction of child molesters; and, in a few cases, assisted law enforcement companies in rescuing sufferers before they may be physically injured. A recent survey of child sexual abuse observed that sharing online snapshots and films documenting crimes, devoted to them reinforces emotions of disgrace, humiliation, vulnerability, etc. [11].

This breaks down the video in critical frames and generates hash codes. PhotoDNA can match the modified image for detection; PhotoDNA Video can detect sexually explicit or child-induced pornography that may appear harmless. Presently automatic equipment like PhotoDNA has become a major tool to fight against online child abuse. It has come up with a great positive difference, technology also can help us to pick out folks who are sexually abusing kids and their photo collections can reason further psychological, emotional, and mental trauma to their victims. These days, PhotoDNA is used by companies around the arena and helps in locating, disrupting, and reporting thousands and thousands of child abuse pics [12].

How does PhotoDNA technology work?

Using an existing PhotoDNA technology solution for images and applying it to key frames of a video as shown in fig-3. Internet Watch Foundation is creating more hashes from videos specifically with the process of PhotoDNA as given below.

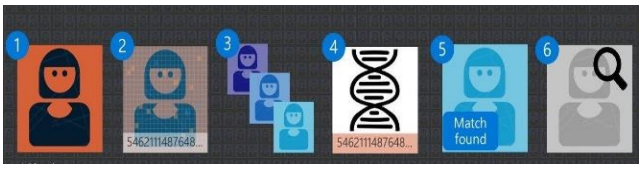

Fig-3: Process of PhotoDNA

Step-1: Child exploitation video tracked and then identified by NCMEC or some other trusted sources.
Step-2: PhotoDNA generates a unique digital sign (a hash) from detected key frames as shown in fig-4.

Step-3: Break up the video to review into key frames and create hashes - use PhotoDNA to reduce the number of frames to check.

Step-4: Hashes of bad images are differentiated with the hashes of the key frames.

Step-5: Matches can then be reviewed manually by agents providing training to handle this content and may only have to view images and not the video.

Step-6: CSAM review process

- Match confirmed:
- Cyber Tip-line Report
- Account suspension
- Content removal
- Create and share all types of hashes

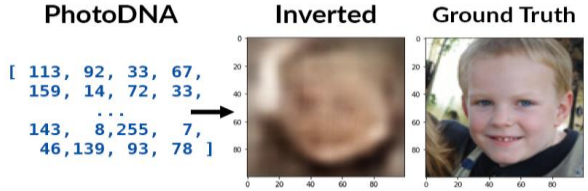

Fig-4: Generated hashes from detected key frames by PhotoDNA

PhotoDNA creates a unique virtual signature (additionally known as a "hash") that compares the signatures of other photographs to obtain copies of the same picture as shown in fig-5. Compared to a website that includes pre-detected illegal hashes of images, PhotoDNA is a top-notch tool to assist those that come across, it is a top-ranking tool to identify and discriminate the abusive items, but this is not a software that can

resize or recognize a photo and items within that and its hash cannot be changed. PhotoDNA is broadly protected in the new visible photos and forensic gear used by law enforcement organizations around the sector.

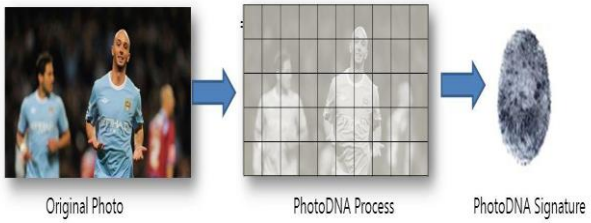

Fig-5: Unique virtual signature created by PhotoDNA

B. Artemis

From social media to SNSs, child molesters are increasingly commonplace and use every available technology under their reach to abuse their potential victims. Preventing child abuse is very critical for any agency that allows person-generated content on its web page. Recognizing the need for cooperation regarding cyber safety, the various most respected generation agencies collaborate to form a coalition called Task Artemis. The result of this multi-year, multi-corporation attempt turned into a textual content-based multidisciplinary analysis software specially designed to deal with child-rearing.

How does Artemis work?

It's an AI system that is predicated on a site of commonplace text patterns based totally on proven -searching for child conversations. Moreover, it includes a danger detection engine, which is designed to evaluate the extent of the danger of a dialogue based totally on content material. The bad guys are mainly exploiting children for sexual purposes using chat activity in multiplayer video games. Artemis works on automatically scanning text-based comments and evaluating them based on algorithms to determine if a user is trying to sexually exploit children. Human moderators will later review flagged comments and determine if a user needs to be reported.

Artemis is a device that scans conversations to find sexual predators on the internet:

The method is used in historic discussions based on textual content. It evaluates and "provides" the features of the conversation and gives a complete measure of possibilities. This measure can then be used as a decision, set by individual companies that use the process when the flagged speech should be sent to human directors for review. Artemis uses a series of icons, the information retrieved by various Artemis icons is then compiled to generate risk identification points. Using risk identification points, a company can determine the likelihood that any discussion will be based on known self-correcting behavior. Based on these possibilities, companies may perform additional reviews in person or perform an automatic action.

There are numerous advantages to Artemis' technique. First, the internet site of commonplace text styles utilized by child abusers turned into compiled data then it uses the facts from many groups and boards. Second, unlike most text-based engines such as Google, Artemis can find research statements through several interactive exchanges. Ultimately, the modular structure of Artemis allows for flexibility and expansion-based requirements of different structures. Further, profile metadata guide and natural language processing algorithms included that have been used within and beyond nameless statistics analysis: one to discover special kinds of speech patterns and the other to discover a way to create a screen. seeing that Artemis cannot get the right of entry to the records associated with the personal account automatically, profile metadata represents extra statistics that are used to help Artemis' essential hazard detection algorithm in accomplishing more correct assessments using textual content communication. Further, by combining this additional information, Artemis is enabled with a function to decide when a person has lied and approximately a guess of his or her age. For an instance, adults who pretend to be kids regularly use the identical emojis and slang as youngsters, but the psychological tactics behind their writing often produce mathematically different textual content. This is because emojis and slang are continually bendy and are not a natural way of writing for human beings of various ages. Using the equal approach, we can also see when one user is created as a behavior mirror for some other people, that's how attackers use it to build relationships and trust.

VI. PROPOSED SYSTEM AND BLOCK DIAGRAM

A. Pre-Processing

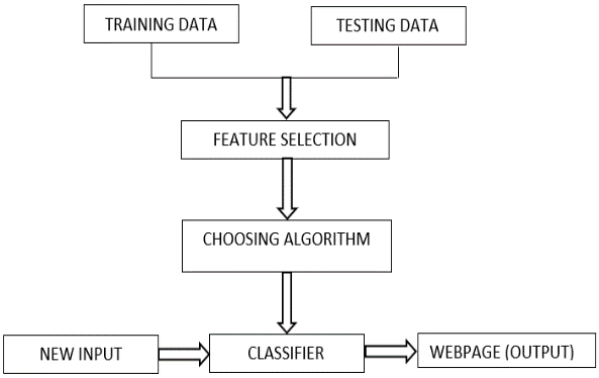

Fig-6: Block Diagram.

Real-world data is usually raw and contains incomplete or missing values which cannot be used for machine learning models, so pre-processing is required to complete the missing pieces of the incomplete values which boosts the efficiency of the machine learning model. For choosing a model, dataset is split into train and test in 3:1 ratio which means Training data having 70 percent and testing data having 30 percent. This split process performed based on train test split model, after splitting we get x-train, x-test and y-train, y-test.

Out of all the algorithms present, the best and most accurate one is chosen after all the algorithms are run and tested against the inputs. It will choose the best possible algorithm based on the type of input and will give the results along with the accuracy. Thus, in our Proposed System we use Machine Learning as a base joined with different calculations to do the cycle.

B. Algorithm

***Support vector classifier*:** Support Vector Classifier or SVC is famous Supervised Learning calculations utilized for Classification and Regression issues.

***Random Forest classifier*:** Arbitrary woodlands are a regulated learning calculation utilized for both characterization and relapse and, furthermore, is the most adaptable and simple to utilize calculation.

***Ada Boost classifier*:** This classifier joins different ineffectively performing classifiers by constructing a high exactness solid classifier.

***Gradient Boosting Classifiers*:** Slope supporting frameworks use choice trees and relapse trees that yield genuine qualities. It shows the likelihood of each info trait for the anticipated state.

TABLE I. CLASSIFICATION REPORT FOR CLASSIFIERS

Support Vector Classifier	**Precision**	**recall**	**f1-score**	**support**
1	0.71	0.90	0.79	30
2	0.81	0.54	0.65	24
Accuracy			0.74	54
Macro average	0.76	0.72	0.72	54
Weighted average	0.76	0.74	0.73	54
Random Forest Classifier	Precision	recall	f1-score	support
1	0.76	0.73	0.75	30
2	0.68	0.71	0.69	24
Accuracy			0.72	54
Macro average	0.72	0.72	0.72	54
Weighted average	0.72	0.72	0.72	54
Ada-Boost Classifier	Precision	recall	f1-score	Support
1	0.75	0.80	0.77	30
2	0.73	0.67	0.70	24
Accuracy			0.74	54
Macro average	0.74	0.73	0.73	54
Weighted average	0.74	0.74	0.74	54
Gradient Boosting Classifier	Precision	recall	f1-score	Support
1	0.79	0.73	0.76	30
2	0.69	0.75	0.72	24
Accuracy			0.74	54
Macro average	0.74	0.74	0.74	54
Weighted average	0.74	0.74	0.74	54

C. Data Collection

In the starting phase, the predators start collecting the data of rich children and start stalking all the activities of children. Then they become online friends of such children and behave in a way that the child starts liking them and all this happens based on the information collected by the predator previously. Now as the predator knows about the likes and dislikes of the child, their hobbies, interests, their family and friends, and many other things, it becomes very easy for them to gain the trust of the child. Once the trust-building process is done with the child, it provides them all the extra information and other stuff when the predators ask them to do so after some time the predator starts blackmailing or abusing them. As per the previous studies researchers used many technologies, they have created the prediction model to reduce the risk and

start training prediction models to train that model we need labeled data.

Once we collected the data or the message content of every user from several conversations, we applied basic pre-processing techniques to the whole corpus, constituting the different ways to identify the similar word and characters. We have combined the content of each attacker with all the victims in the conversation to see the top vocabulary words used by both groups of users. Moving forward we can also identify the terminology used by the abusers or the predators [13]. We found that words including “like”, “want”, “know”, “call”, “home”, etc. are the frequently used terminology of the abuser and the abused. We noticed that these words belong to different word categories including approach words (eg. “meet”, “together”, “car” “room”, “hotel”), family words (e.g. “mom”, “dad”, “sister”, “brother”), and relationship words (e.g. “boyfriend”, “partner”, “date”). In addition to it, we also saw other words such as "pretty", "beautiful", "cutie", "sweetie", “princess”, “like”, were used most of the time, but these types of words used by predators so we can track predators, victim and the chat between them with the help of this data easily. The collected data plays a very vital role because it is the beginning or starting phase to gather information before performing any action on the predators, the collected data must be in an understandable format so that further processing will be done in an efficient and fast manner.

D. Analysis of Data

After data collection we move ahead with data analysis, in the analysis part we operate on the data with the help of big data analysis and after that, we can categorize the data in such a manner that we can easily fetch the data whenever it is required and it is a very helpful part of the analysis, for example, we categorize data into male/female, age, locality, harassment factor, etc. In addition to it, we also saw other words which are being frequently used by both groups. These things can be analyzed in data analytics. The result of this kind of analysis can help the analyst to come to the right conclusion.

Once the data is pre-processed, the next step in machine learning is to derive the optimal feature vector from the chat. This content can provide the data that can differentiate the terms used by both the abuser and the child [5]. On behalf of analyzed data, we came to know that there is a certain word that is mostly used by the child and there is a different set of words that are mostly used by the predator but on the other hand, there are a set of certain words that are used by the both, these words are categorized under eight categories according to Mood Book dictionary [5] further explains this. The dictionary categorizes these words and connects them with emotions which are categories for emotions such as fear, anger, sadness, joy, surprise, disgust, optimism, and anticipation. For example, the joy category includes words like “happy” and “awesome” and the sad category includes terms like “cry” and “hopeless”. This dictionary also lists these words as having positive and negative traits and connects them with eleven emotions. The first ten user attributes of i (Ei1, Ei2, ..., Ei10) correspond to the eight senses and two (positive and negative) emotions of the Mood book categories, respectively. Later on, the authors [14] found words such as “lips”, “legs”, “mouth”, “eyes”, “tongue”, and “hair” and categorized them as body-parts words. On behalf of analyzed data the terms can be categorized as abuser’s vocabulary or a child’s vocabulary and the difference between the both is visible For example, words like “sweetie”, “feeling”, “please”, “body”, “touch”, “18” are mostly used in abuser’s messages whereas “kno” (know), “2nite” (tonight), “broke”, “hurts”, “idk” (I don’t know), “14” are generally used by children in their message in the same contest the word horny is used by twice only out 82 time by the child so this word can be put under the category used by the abuser most of the time). Category Word(s) Example (from dataset) P: predator, A: abuser, V: victim Approach words apartment Come to my apartment. Here I am alone. Will u come? (P) Connection words chatting ... feeling better now as I’m talking to you (V) so nice, awww u r amazing, awww ur so cute, etc. (V) words horny ... I am getting horny while talking to u. (P)

E. Implementation

Given informational collection is partitioned into testing and preparing information to keep away from the complicity of the calculation execution. In the preparation set, a realized result is available and this information is learned by the

module to sum up different information later on. The test dataset or subset is utilized to test the model's expectation on the given subset. It undergoes pre-handling to improve the dataset clear space and so forth. For pre handling, different innovations like little max scalar, standard scalar, etc. are used. Yet, we utilized little max scalar as it were. And afterwards going for highlight extraction in the informational index, every one of the various qualities are elements of individual. Highlight extraction only gets the qualities from the given dataset. After that, we are playing out the calculation for expectation. Each element in a given reach is changed. Each component is exclusively scaled and deciphered to such an extent that it is in the given reach on the preparation set. For example, somewhere in the range of nothing and one. One more comparative and elective way to deal with

Z-score standardization (or normalization) is the Min-Max scaling or "Standardization". Information is scaled from 0 to 1 in this methodology [15]. Nonetheless, the expense of having this specific limited range - in similitude to normalization - is that we will get more modest standard deviations, which stifles the impact of the exceptions.

A Min-Max scaling is done through this equation: $Xsc = (X - Xmin)/(Xmax - Xmin)$.

F. *Experimental Results*:

For all sets, each image sampled from the Tiny Image net dataset or the LFW dataset constitutes as a separate class, and all the transformations of that image belong in that class [16].

TABLE II. CHARACTERIZATION AND DESCRIPTION OF CHILD

Attributes	**Description**
Age	Age in Years
Sex	Patients gender (Male, Female)
Locality/Region	North, South etc.
Abuse Type	Sexual abuse, physical abuse, emotional abuse and neglect.
Recognizing Signs of Child Abuse/Harassment factors	Mental health issues, increased stress and anxiety, depression, acting out violently, and low self-esteem
Data type	Image, text, video
Results	As per the data type

TABLE III. THE STATISTICS FOR DATASET

		Training		**Testing**		
Data-set	**Classes (Original Images)**	**Trans-formations**	**Total**	**Classes (Original Images)**	**Trans-formations**	**Total**
Cat-1	100500	44	4420000	10050	44	442000
Cat-2	10050	44	420000	1005	44	44200
Cat-3	10050	9	90450	1005	9	9045
Cat-4	450	6	2700	100	52	5200

VII. DISCUSSION AND CONCLUSION

The main motive of this research work is to enquire about the importance of ML and AI technology with the combination of GIS and data mining in preventing long-term losses associated with online child exploitation. PhotoDNA and Artemis play a vital role and these technologies are emerging technologies. PhotoDNA helps us in image matching, and we can track one's image in a fast and effective manner Artemis, is an artificial intelligence device that relies on a database of common textual content styles based totally on tested child and predator conversations. It consists of a danger detection engine, designed to assess the level of danger of a discussion based on content material. In conclusion, our results show that our proposed approach performs very accurately for image fingerprinting. We noticed that deep learning techniques in general are significantly more accurate than image processing based techniques for all transformations. Hence, these insights suggest that as deep learning is getting more prevalent, there is a critical need to move to deep learning for applications of security as well. We believe that our algorithm's performance can effectively solve the problem of distribution of prohibited digital images on the internet, and similar techniques should be urgently adopted in the industry in place of the current image processing based algorithms.

The overall conclusion throughout the process of children trafficking and exploitation is a major cause, and our proposed solution is very helpful to reduce the exploitation. We aim to develop this research model physically tested. We

will be able to do some test cases in the future and after that, we can give this creative idea to people or to legal offices to track the trafficking of children. This project has less cost which is a major feature of this project, and everyone can easily get this project as it is user-friendly. Stopping child exploitation has usually been critical for everyone and we're honored to conduct a study on this challenge. Over the years, it has been furnished with essential sources for the information and identification of predators. Even though we've made a significant study, the era keeps on emerging, and we are constantly striving to afloat childhood and help children to grow in a healthy and safe world.

ACKNOWLEDGMENT

I am grateful to all of those with whom I have had the pleasure to work during this and other related projects. Jamia Hamdard has provided me extensive personal and professional guidance and taught me a great deal about both scientific research and life in general.

REFERENCES

[1] Siregar, Mulia, et al. "Online sexual exploitation as Globalization Homemade Problems." (2020).
[2] Dutta, Nitul, et al. "Introduction to Digital Forensics." *Cyber Security: Issues and Current Trends*. Springer, Singapore, 2022. 71-100.
[3] Pereira, Mayana, et al. "Metadata-based detection of child sexual abuse material." *arXiv preprint arXiv:2010.02387* (2020).
[4] Vogt, Matthias, Ulf Leser, and Alan Akbik. "Early Detection of Sexual Predators in Chats." *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. 2021.
[5] Ngejane, Cynthia H., et al. "Digital forensics supported by machine learning for the detection of online sexual predatory chats." *Forensic science international: Digital investigation* 36 (2021): 301109.
[6] Ramiro, Laurie S., et al. "Online child sexual exploitation and abuse: A community diagnosis using the social norms theory." *Child abuse & neglect* 96 (2019): 104080.
[7] Wani, Mudasir Ahmad, Nancy Agarwal, and Patrick Bours. "Sexual-predator Detection System based on Social Behavior Biometric (SSB) Features." *Procedia Computer Science* 189 (2021): 116-127.
[8] Knox, Shawn, et al. "What's really 'Happning'? A forensic analysis of Android and iOS Happn dating apps." *Computers & security* 94 (2020): 101833.
[9] Benavente, Beatriz Teresa, et al. "Risk factors for commercial sexual exploitation of children and adolescents: results of an international Delphi panel." *Psicothema* (2021).
[10] Açar, Kemal Veli. "Organizational aspect of the global fight against online child sexual abuse." *Global Policy* 8.2 (2017): 259-262.
[11] Dushi, Desara. "Combating the live-streaming of child sexual abuse and sexual exploitation: A need for new legislation." *Second international handbook of Internet research* (2020): 201-223.
[12] Açar, Kemal Veli. "Framework for a single global repository of child abuse materials." *Global Policy* 11.1 (2020): 178-190.
[13] Bond, Emma, and Andy Phippen. "Tackling Teen Sexting—Policing Challenges When Society and Technology Outpace Legislation." *Policing in the Era of AI and Smart Societies*. Springer, Cham, 2020. 157-177.
[14] Agarwal, Nancy, et al. "Predatory Conversation Detection Using Transfer Learning Approach." *International Conference on Machine Learning, Optimization, and Data Science*. Springer, Cham, 2021.
[15] Al-Nabki, Mhd Wesam, et al. "Short Text Classification Approach to Identify Child Sexual Exploitation Material." *arXiv preprint arXiv:2011.01113* (2020).
[16] Özçalık, Cennet Kara, and Rahime Atakoğlu. "Online child sexual abuse: Prevalence and characteristics of the victims and offenders." *Journal of Psychiatric Nursing* 12.1 (2021): 76.