Cybercrime Prevention on Social Media

Balkrishna Shah, Nitu Sharma, Saloni Bandgar
UG Student, Dept. of Information Technology
Vidyavardhini's College of Engineering and Technology,
Vasai, India

Prof. Sainath Patil
Assistant Professor, Dept. of Information Technology
Vidyavardhini's College of Engineering and Technology,
Vasai, India

***Abstract*— Since in the early years of the 21st century, crimes tend to be committed away from the eyes of the majority of society. The advent of social media in the past decade has led to a new sort of 'performance' crimes, where people create accounts of their law-breaking through text, images, and video, which are then digitally distributed to the public on a large scale. This leads to increasing social crimes day by day. Crimes that are observed all over the globe are "posting nude photos", "sharing abandon posts ", "cyberbullying", sending bad or vulgar words in personal chats which is a part of cyberharassment". The primary objective of this project is to sight such cases over social media and execute some inhibitory course of action in order to avoid above stated incidents. Our social media could conclude the surprising, unpleasant, and unwanted content that could be abusive-disrespectful text or nude pictures, or unpermitted actions and facilitate users removing it or preventing it from unfolding everywhere. This may facilitate the general public to assure safety and nonviolence. It also includes an intrusion detection system which makes the platform more secure using anomaly-based detection by deep learning**

Keywords*— Cybercrime, cyberbullying, social media, intrusion detection, machine learning, neural networks.

I. INTRODUCTION

Social networking platforms are being widely used today for multiple purposes like entertainment, networking, etc., and becoming a boon for everyone but on another side, with the increasing number of users on social media leads to a new way of cybercrimes. Cyberbullying is becoming a major issue and is defined as an intentional or an aggressive act that is carried out by a person or groups of individuals using repeated communication overtime against a victim who cannot easily defend him or herself. With the inception of the internet, it was only a matter of time until bullies found their way onto this new and opportunistic platform. Using services like instant messenger, and other social media platforms bullies became able to do their nasty deeds with anonymity and the great distance between them and their targets. According to the Cambridge dictionary, the term cyberbullying is defined as the activity of using the internet to harm or threaten another person, especially by sending them unpleasant messages.

The main factor that separates cyberbullying from traditional bullying is the effect that it has on the victim.

The main issues that we have covered are as follows:

Cyberbullying is a way of threatening or intimidating a person either through messages or by posting objectionable content on the internet. This activity has caused a lot of damage to users of social media networks, especially amongst the youth. Several cases have been reported which show that Cyberbullying has caused teenagers to go into depression and may commit suicide.

Intrusion: Cyber-attack incidents are increasing with the increasing use of the internet. Cyber-attack is the virtual life of bullying in normal life. In this attack, a person encounters such situations as harassment, threats, and blackmail. The attack may be in the form of the capturing of the persons' passwords or giving psychological pressure.

Given the consequences of cyberbullying on victims, it is urgently needed to find a proper action to detect and hence to prevent it.

As social media platforms are vulnerable to such crimes, our solution provides a social media platform that will be more secure and able to take preventive measures against culprits to build a safe social environment.

A. Problem Statement

With the exponential increase of social media users, cyberbullying has emerged as a form of bullying through electronic messages.

Cyberbullying involves posting and sharing wrong, private, negative, harmful information about the victim. In today's digital world we see many such instances where a particular person is targeted.

For avoiding issues related to hacking and non-repudiation, it is necessary to introduce some preventive measures like an intrusion detection system

Given the consequences of cyberbullying on victims, it is necessary to find suitable actions to detect and prevent it.

II. RELATED WORK

Cyber safety is the foremost problem for all the security giants in the world. Recently there have been many kinds of research done to make the internet a secure place.

A. Social Media Cyberbullying Detection using Machine Learning

An idea in this literature consist of three main steps are Preprocessing, features extraction, and classification.[1]

- **Preprocessing**: In the preprocessing step we clean the data by removing the noise and unnecessary text.
- **Feature Extraction**: The second step is the features extraction step. In this step, the textual data is transformed into a suitable format applicable to feed into machine learning algorithms. TFIDF and sentiment analysis is used in this step.
- **Classification**: The last step is the classification step where the extracted features are fed into a

classification algorithm to train, and test the classifier and hence use it in the prediction phase.

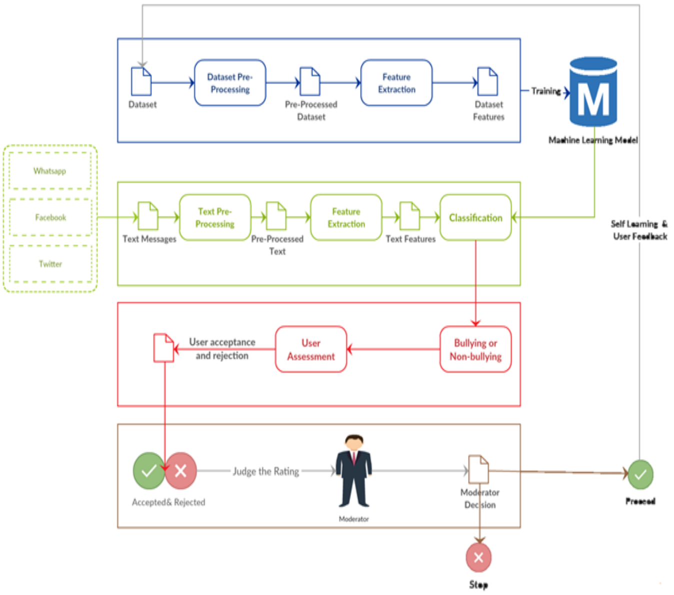

Fig. 1. Social media Cyberbullying detection

Author has used two classifiers, namely, SVM and Neural Network. Generally, the evaluation of classifiers is done using several evaluation matrices depends on the confusion matrix. Among those criteria are Accuracy, precision, recall, and f-score.[1]

Accuracy = TP+TN/TP+TN+FP+FN

Precision = TP/TP+FP

Recall = TP/TP+FN

F - Score = 2precisionrecall/precision+recall

They have used a cyberbullying dataset from Kaggle which was collected and labeled by the author Kelly Reynolds et al.in their paper. The comparison of SVM and NN in terms of Accuracy is a follow-

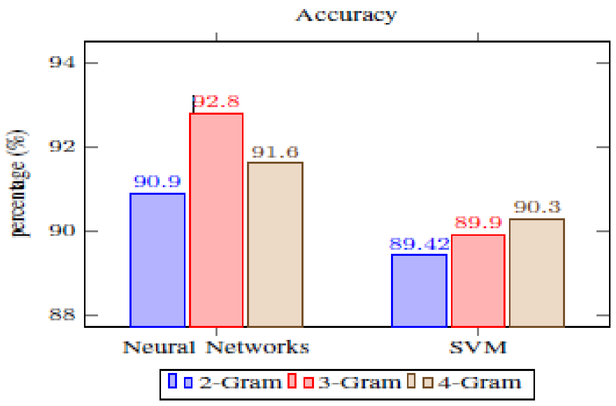

Fig. 2. Comparison between SVM and Neural Network in Terms of Accuracy

B. Detection and Prevention measures for Cyberbullying and Online Grooming

In this system, authors have developed a social media platform like Facebook where users can post messages and images. This system will detect adult images and improper comments or other messages and restricts from posting them.

Here they have used two datasets where the User Messages are classified using Bad Words Dataset and Sensitive Words Dataset and the user's bad count is recorded into the database. If any user exceeds the limit of abusive posts or adult images then he/she is automatically banned from using this social networking portal. And hence, this portal is more secure and reliable than Facebook.

Illegal posts have been checked using Sentiment Analysis based on text. Sentiment analysis is done on user posts and comments using NLP (Natural Language Processing) algorithms to detect user sentiments.[2]

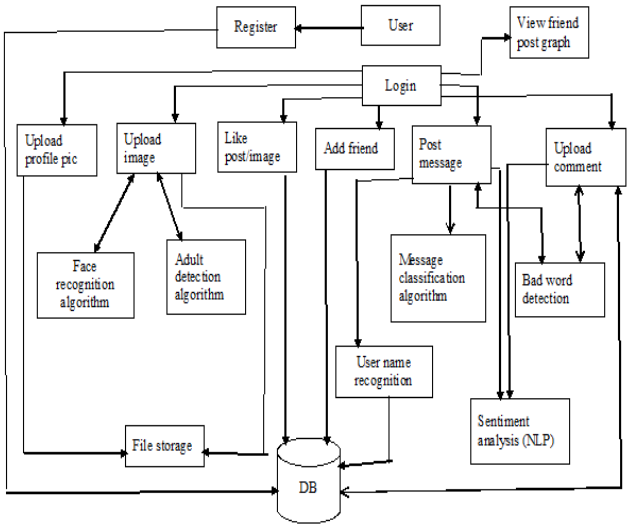

Fig. 3. Detection and prevention of cyberbullying

Algorithms used are-

Adult Image Detection:

Any user who uploads an adult image is warned about posting such images. Every user has few chances before their account is blocked. If the user is found to be posting such content, there is a provision for the entry in the database which checks if the account shall be allowed to continue or not.

Irrelevant Posts Detection Algorithm:

This algorithm extracts keywords from text files containing different categories of data and user messages are scanned to find whether they contain those keywords to classify messages into the Crime/Worst/Riots category and to find the sentiment of a message.

NLP Algorithm:

The analysis of comments and the amount of positive or negative reactions combined with sentiment analysis through different text mining modules give a rough overview of a user's social status. User Post is parsed into

Tree Structure and then words of post are analyzed from tree structure to find the sentiment of the post. Facebook posts from the user's account.

C. Deep Learning in Intrusion Detection Systems

For Intrusion Detection systems; many techniques have been developed for modeling the data and create tables by classifying the modeled data.

- **Statistical:** By examining the user and system behavior, a statistical model is created. Some of the statistical methods used in intrusion detection are Principal Component Analysis (PCA), Chi-square distribution, Gaussian Mixture Distribution.
- **Artificial Neural Networks:** It is used to examine and learn the behavior of data in the system. With an enhanced form of ANN, some authors prefer to use Deep Learning for its efficiency
- **Support Vector Machines:** It is the most preferred method for intrusion detection systems. Support Vector Machines distinguish between data from two classes in the most appropriate way with a feature vector.
- **Data Mining:** It is used to extract patterns by finding the relationship between data and users.
- **Rule-Based Systems:** It is developed by people who specialize in a specific area. These people examine the system traffic and form rules and attack detection is done in this way.

IDSs can be classified according to their techniques; Signature-Based and Anomaly-Based.[3]

- **Host-Based IDS:** server tries to detect attacks by listening to the traffic, registration files, and transactions.
- **Network-Based IDS:** listening to all the traffic directed to the network, recording the content of each data packet passing through the network, cutting off attacks when necessary, and creating reports.
- **Signature-Based IDS:** is used to detect known attack types.
- **Anomaly-Based IDS:** is used to detect unseen attacks.

III. COMBINED STUDY

From combined literature review some results have been found. Those are as follows:

Authors /Techniques	Identification Accuracy
SVM For Sentiment Analysis	92.4%
Neural Networks	91.7%
Naïve Bayes classifier, decision tree, SVM for text analysis	66%
Color-based skin detection	89%
Anomaly-based detection	75%

TABLE I. Combined Study

The results in Table I shows that every method has a certain accuracy according to the related work papers. So, this work can also consider these methods to get better accuracy or accuracy more than these. This paper is using various combinations of methods to achieve higher accuracy.

IV. METHODOLOGY

Till now the papers and research work that we've referred to have been working on some particular parts of our project. We will be working collaboratively on all the parts viz. extraction of data, detection of bullying activities, and hence protecting the users from becoming a victim.

This project includes a more secure platform for social networking by the intrusion detection system. A further addition to this project is that we would be taking actions against the culprits so that the people who come with an intention of committing a crime would not be given many chances. The continuity pattern will be checked for posting unwanted and unpleasant content. After a limit of attempts first, the user will be warned.

The ones who are found to be violating the rules will be permanently banned from the portal after a limit of attempts till further investigation.

A. Working

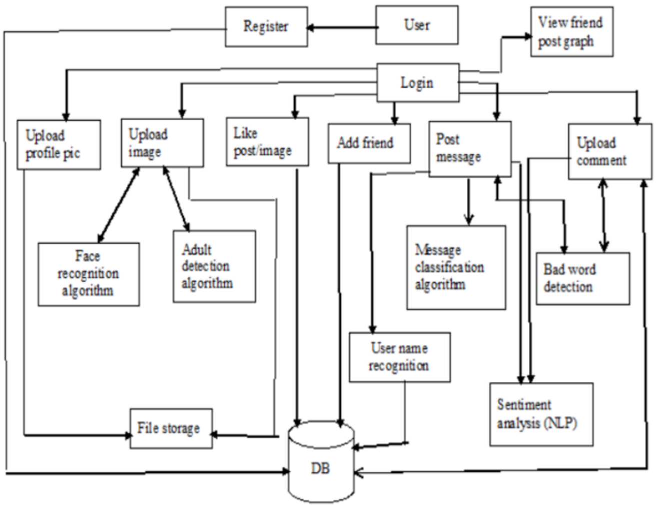

Fig. 4. Proposed System

To implement a system that will prevent cybercrimes such as cyberbullying and intrusions, social media platform similar to Facebook will be developed.

- In this platform, users will be able to register themselves and perform various tasks like:
  - Add friends
  - Upload Profile Photo
  - Upload Posts
  - View friend's Posts
  - Like other's Post or images
  - Comment on someone's Post
  - User will be able to send messages to their friends
- When any user uploads any post, it will be passed through the Post classifier to divide the post into illegal posts or adult image posts.

- Illegal post detection algorithm will check if any violence-based objects are present or not.
- Image-related posts will be passed through an adult image detection algorithm whether nude images and unpleasant images will be checked.
- When any user comments on any post or sends messages to any friend or other user then sentiment analysis is done to find the intention of the message and get the amount of positive or negative reactions on the messages to check whether the user is trying to harass somebody.
- The frequency count of Vulgar words and bad sentiments will be stored in the database that how many times the user is doing such actions
- Continuity pattern will be checked through the frequency count
- A particular limit will be set for the frequency pattern.
- If the user exceeds the limit of sending such illegal posts or comments and messages which are having a bad impact on other users will be warned for the first time
- If the person still continues to do so then the user’s account will be reported and blocked until further investigation takes place.
- If somebody posts something mistakenly, then the system provides a feature to delete the post.
- For avoiding intrusions and non-repudiation our solution also proposes the intrusion detection system.

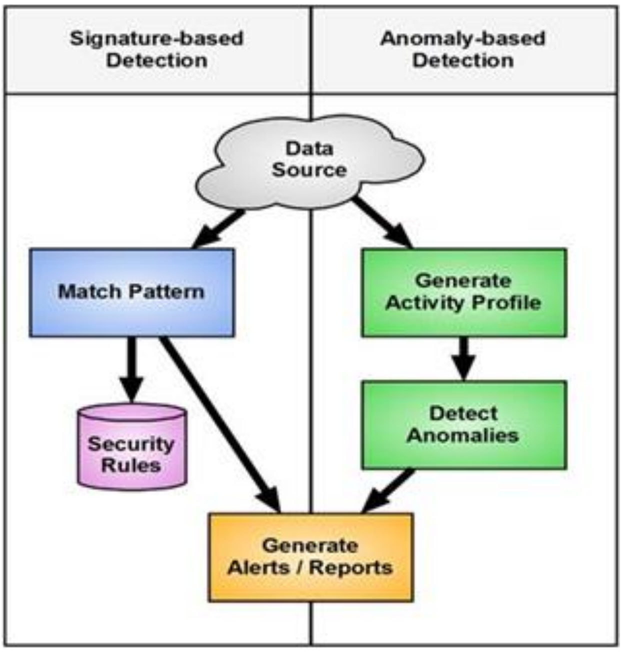

Fig. 5. Comparison between Signature-based and Anomaly-based IDS

- The papers we have referred to involve Signature Based IDS
- IDS Signature detection works well with the threads that are already determined or known.

Drawbacks of Signature-based IDS: -

- Database of only attack pattern
- Detect Only Known attack
- Cannot Identify new attack

Our approach uses Anomaly-based intrusion detection which will check the behavioral pattern of the user.
And perform behavioral analysis to detect anomalous behavior to avoid intrusions.
Hence, providing a safe and secure social media platform.

V. Algorithm Details

A. Adult Image Detection Algorithm

The Nudity Detection Algorithm is predicated mostly on observations that generally, nude photographs contain large quantities of skin, humans have different skin tones, and skin areas in nude snapshots are enormously close to every other.

These observations require the identity of skin and skin pixels in a photo to be classified. The diagnosed skin pixels are analyzed to workout which among them are related or form continuous regions.

These skin areas are in addition analyzed for clues of nudity or non-nudity.

In general, the Nudity Detection Algorithm is composed of the subsequent steps:

1. Detect skin tone pixels within the image.
2. Locate or shape skin regions based on the detected skin pixels.
3. Analyze skin regions for clues of nudity or non-nudity.
4. Classify the post as nude or not.

The first step of the algorithm is to identify the amount of skin pixels in the Image. Since the image is the color image and skin detection function classify each color pixels into the skin.
It gives: -

0: input pixel is not a skin-pixel

1: input pixel is skin-pixel.

If the amount of skin pixel is more than 75% then it goes to the second step.

B. Skin detection techniques

Involves the formulation of an efficient mathematical model to represent the skin color distribution.

- Range-based
  - transform an input pixel into a proper color space
  - Skin detection classifier to label the pixel whether it is a skin or non-skin pixel.
- Histogram back-projection
  - The histogram is a spectrum of intensity repartition.
  - A list that contains the number of pixels for each possible value of the pixel.

C. Illegal Post Checking Algorithm

Illegal posts must be checked using Sentiment Analysis based on text.
Sentiment analysis is done on user text-based posts and comments using Natural Language Processing algorithms to detect user sentiments.

Object detection algorithm is used to check any violence-based objects is present or not in the user's post

D. Natural Language Processing (NLP)

The analysis of comments and the amount of positive or negative reactions combined with sentiment analysis through different text mining modules give a rough view of a user's social status. User message or comments is parsed into Tree Structure and then words of post are analyzed from tree structure to find the sentiment of the message or comments.

In addition to image analysis, text analysis of the textual content of shared images can provide helpful information in order to improve classification accuracy. Another problem analyst is confronted with, is the distinction between pornographic and child pornographic images, the distinction between legal and illegal content. The main perspective of the system is accessing the posts from the user's account.

E. User Behavior Detection Algorithm

It is used for security purposes for analyzing the behavior of users.

It triggers or performs an action when there is a drastic change in user behavior observed. It maintains a journal of the normal performance of the user. In turn, the system detects any anomalous behavior for a particular user when there are deviations from its normal behavior or patterns.

For example, if a user never Logged in for 3 months then suddenly his/her activities increase on social media then there is called pattern change or behavior change.

Neural networks are used for designing this algorithm to observe behavioral patterns optimistically.

VI. RESULTS AND DISCUSSION

Our social media designed in this work is built on Django backend and the Algorithms used in this paper are based on Convolutional Neural Network (Which can take an input image and be able to differentiate between one from other as trained) which leads to high accuracy and desired output. We present the two training methods for adult image detection i.e., Image classification followed by Object Detection.

This designed social media can warn the user for posting bad comments on public posts and if its limit exists then his/her account will be suspended and further investigated. This system is able to restrict nude posts from posting on social media that is also a new feature to the world on social media. We present user behavior detection which is a new concept that analyzes user behavior and feeds it to Neural Network to predicts exponential changes in daily user routine behavior.

VII. CONCLUSION AND FUTURE SCOPE

In this paper, we have developed a comprehensive product and are attempting to set up a system to protect the young generation from cyberbullying attacks. We have implemented our own social media and were able to design most of the features that are discussed in this paper. Actions are also performed on social media if rules are violated. We have tried to show a real-time scenario for all the features mentioned in this paper.

Adding multilinguality to our project is our future application and we will try to provide our feature to existing social media by migrating to API concept or by providing filtering features with chrome extensions.

REFERENCES

[1] John Hani, Mohamed Nashaat, Mostafa Ahmed, Zeyad Emad, and Eslam Amer. "Social Media Cyberbullying Detection using Machine Learning International Journal of Advanced Computer Science and Applications, (IJACSA) Vol. 10, No. 5, 2019

[2] Aishwarya Upadhyay, Arunesh, Akshay Chaudhari, Sarita Ghale, Prof. S. S. Pawar "Detection and Prevention measures for Cyberbullying and Online Grooming" International Conference on Inventive Systems and Control (ICISC-2017)

[3] Gozde Karatas, Onder Demir, Mostafa AhmedOzgur Koray Sahingoz" Deep Learning in Intrusion Detection Systems" International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism Ankara, Turkey, 3-4 Dec 2018