



INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

The use of Machine Learning to detect obscene contents on social media.

1Muhammad Abul Kalam, 2Dr R Udayakumar

1Research Scholar, 2Professor

1Bharath Institute of Higher Education and Research,

2Bharath Institute of Higher Education and Research

Abstract—In this paper I am writing about the different approaches to find the fake content, abusive content, vulgar content, immoral content and any content that are against the regulatories of the Indian IT act and how to delete those content for clean and safe use of the social media. This paper contains the different type of social media misuse for the personal benefits or to defame and tarnish and individuals image or business and also to disturb the peace and harmony of the people of this country and this paper also contains the need for regulation of internet so that our country remains safe from the cyber attack launched from other countries and within. This paper also contains the different steps taken by the government to control misuse of social media and social media crime and also presenting the possible solution to trace and block and delete any illegal contents, illegal websites, unauthorized url etc. This paper have both manual and automatic strategy to be taken up to control the unauthorized use of social media and the misuse of social media.

Keywords— Internet, Cyber crime, social media, ethics, cyber law.

I. Introduction on social media.

Social media means the communication between human directly or indirectly on the electronics device through internet. The most popular social media platforms are Facebook, Whatsapp, Twitter, Youtube, Instagram, LinkedIn etc. And social media has become very much a part of our life that we started socializing on social media virtually instead of physically. Social media has both positive and negative effects. So it is upto the user and also upto the rule of that particular country in what way the social media should be used.

As we know that social media has been very popular from the time internet has been widely available at an affordable cost and the availability of smart phone at low cost. Everyone from children to adult to old, are using the social media on a regular basis to share their opinions, to share anything that they feel interesting. Social media has become like a big family of this world. People have been connected so widely that in social media we get everything virtually. Through social media anyone from one corner of the world to someone in another corner of the world are

connected. Strangers have become friends and users can learn many many new things available in this world. In social media, we get diverse information ranges from ones personal information to celebrities and local and international news, entertainment, daily updates, current affairs, information about celebrities, politicians and businessman etc. We get update of education, weather, sports, trip, invention, discoveries, stoke market, electronics devices, war, and everything that we get anywhere in this world virtually. At the same time, people have also shared so many contents that are not acceptable to all. Social media is very useful to gain knowledge and to know many new and past things but at the same time social media is dangerous because many people shared too much personal and many people have started taking social media too much serious. The thing is that we should use social media within a limit with ethics. All are different and so sometimes users become very vulnerable to strangers and sometimes even lead to suicide and depression and increase in cyber crime case. There has been many news came in the past about the misuse of social media like stalking, sharing obscene contents and using abusive language to express their sentiment. As we all know that internet has no boundary and internet is totally virtual and there is no any particular limit. Number of users have been increased very fast after the availability of mobile phones and mobile data at a reasonable price. No one hesitates from signing up in the social media. Most social networking site does not ask identity before creating an account and the server of many of the social networking are available in some other country also. And so it is very difficult to trace the user and also their location. Anyone can create account either using their true name or by using some other identity. This is one of the biggest problem. And even internet user can hide their ip address using proxy address. Social media is quite useful because we can use it to learn new things, to keep our opinion on something and also we can express our likes and dislikes and also we can discuss on any issue. Social media can also be used to promote business, to share any new invention and new findings. Social media has made it possible to expose the talent that otherwise would be impossible to bring into this world. We can know about what is happening where in this world. We can also make the use of social media to make different type of campaign. We can use social media to expose the good and bad deeds. Social media is also quite useful in exposing someones wrongdoing and also in giving justice to victimized people. Social media has been used to exposed both good and bad. Social media is quite useful for one and all. But there is also the other side of social media. Users can misuse social media. Before there was social media, it took longer to spread anything but now, because of social media it takes seconds to spread a news from one person to another to others. Any user can post any content (text, image, video, speech, etc.) from anywhere on social media platform. Typically users of these platforms are identified by the profiles they create.

In the social media, users share so many personal details starting from pictures to the place they visit to their exam or interview results to details like marriage, love, sex, caste, religion and what they do and don't do on a regular basis. The users have shared so many personal things that it is like someone knows everything about some others. So users have become very vulnerable to any type of target. And some people have even taken negative advantages like sharing contents to provoke a person, religion, caste, race, countries or someones belief etc.

It is very difficult to make every users aware of all the possible danger of sharing too much details. Out of the all users, more than 55% are below 25 years of age and the people of the age group 25 to 35 years are only 30% and the remaining 15% users are the people of the age group 40 years to 75 years. We see here that maximum of the users are people of the lower age group and they are quite immature and their understanding and self control power is quite less and so

these are the people who are very vulnerable to any type of cyber bullying and many cyber crime that ranges from scolding, abusing, share obscene contents of the self and others and also making obscene and vulgar comment and which are very offensive and against the ethics and moral of social media use.

There has been many cases of cyber bullying, many cases of sharing obscene pictures and many cases of sharing unacceptable content and lakhs of abusive and offensive comments.

Many users share their good and bad personal relationship status and due to uncontrol emotions, sometimes they started sharing the private conversations and pictures shared during their good times and this leads to cyber crime and criminal proceeding and sometimes even leads to suicide, sometimes some users share the poor exam results of their friends and make a mass bullying which sometimes lead to cyber crime plus suicide of the victim due to insult and loss of self respect.

Sometimes some secretly recorded photos and videos are shared and which lead to cyber crime and mass sharing and suicide and criminal proceedings.

Social media can also be used to blackmailing, abusing, circulating fake news to defame others. So we should search a way to detect and delete such derogatory content so that the cyber related crime reduces to the maximum possible.

Fake news is a type of propaganda that consists of eliberate misinformation or hoaxes. It can include text, visual, audio, data report etc. Fake news is written and published with the intent to mislead in order to damage an agency, entity, or person, and/or gain financially or politically, create disturbance and unrest, often using sensational, dishonest, or outright fabricated information to increase readership, online sharing, and Internet click revenue. The typical attributes of fake news are that it spreads fast, is doctored, is incorrect, is malafide, is intentional, is sensational/glamorous, is unverified, and goes viral.

II. Social media advantages:

- Education:

Social media can be used for educational purpose. Users can make group and discuss any mathematical or engineering problems and also can involved in sharing their new solution to some problems and how they did it etc. And in leisure time, people can come to the group to discuss so many useful educational things.

- Updates and information:

Latest update related to education, science and technology, medicines and invention, what is happening where etc can be known from the social media.

- Connectivity:

All forms of people from across the globe irrespective of their belongings and background connected through the social media and can learn new culture, food habit, believes and living styles etc,

- Help:

Maximum of the global population are available in the social media and it is easy to ask for any type of help. May be help for treatment, help for running a free educational society, help for donation of child care home, help for justice etc.

- Guides government and agencies to fight crime:

Social media is useful even in tracing criminals from their accounts. Criminals log into their account and that left the information about where about they are and also law enforcing agencies can take the help of social media users to trace criminal.

- Improvement in business reputation:

Any business agencies can share details about their products in the social media and the same can be shared by multiple users and it finds foothold of their business.

- Building communities:

Many groups are formed and many national and international events can be organized which bring people from different places and help in building big communities.

III. Social Media Misuse:

When social media is misused, it leads to sexual harassment, criminal offense, employment obligations breach, unlawful discernment, misconduct, violation of students, staffs or parent's privacy, legal liability exposure, etc. These are illegal as well as serious activity that takes place due to social media misuse.

Some misuses that are associated with the social media are:

- Defamation:

Defamation can be fake or real. Defamation means associating something with a person or group of people which can make others distancing from that person or group, which will make other hating or avoiding that person or group, which make the name of that person or group become very dirty and bad in front of their own society and also in front of all others. Defamation is a heinous crime may even lead to death. Defamation is done mostly against a politician, celebrities etc to tarnish their image. The individual or group of individual can file a cyber crime case against the involved and necessary actions will be taken up as according to the Indian IT act. We can share defamation message through facebook, twitter, whatsapp etc.

- Copyright:

Copyright is a right to protect the creativity of a person or a company. Copyright protects music, lyrics, scripts, books, arts, science, movies, songs, performances etc. copyright helps creators with the legal right to control the use of their creations. Using any of the copyrighted contents without the proper written permission of the creator will lead to criminal case under the Copyright Act 1968.

- False and unreliable information:

One is hacking and the other one is creation of fake account. By creating fake account, false information are being shared which leads to multiple problems. And sometimes hacking the account and intrude into others private data and this leads to fight, jail, death etc.

- Sexual Predators:

It is easy to make someone fall to someones else love etc in social media. And also there were many cases of making people prey on false promise of job and modeling. Anyone can use the pictures of celebrities or some other people and can attract others.

- Cyberbullying:

Many use social media to bully their friends etc. Some friends make fun of their cool and innocent friends on social media by making bad and provocative and insulting comments and also by sharing some secret which hurts the sentiment of the person.

- Loss of Privacy:

Many people shared and reveal many private information that they shared during their good times. Sometimes employee shared and reveal private information about their employers and the organization in anger as a vengeance .

- Fraud:

Nowadays the most misuse of social media is fraud. Fraud can be in the form of winning lotteries, in the form of job, in the form of giving modeling consignment, in the form of huger offers etc.

- Child pornography:

Child pornography means indulging anyone under the age of 18 in any sexual activities. This is one among the many cases which may make the person punish with longer period of time under the bar.

- Stalking:

It means harassing any individual by sending message or by calling or sharing something insulting or shameful on the social network.

- Time depletion:

Excessive use of social media and the addiction leads to bad performance in exam as well as in daily activities and this is very harmful to the life of students and employees.

- Lack of Social skill development:

Many people have started believing social media as their real life and spending too much of their time in interacting with unknown people and also started sharing their personal life too much from love failure to marriage problem to job loss and depression etc. And many even involved in online love and sex etc.

- Drug and alcohol addiction:

Many drug and alcohol addicts share their experience on social media which attract many just in the name of style or show off and this leads to attracting many more into drug addiction. And to fulfill their needs many even started involving in criminal activities.

- Deaths due to social media:

Many people have started sharing their daredevils stunts and guts and the same thing copied by others and sometimes leads to death.

- Affects self-esteem:

As we know many are addicted to the social media. Less sleep and more on social media. Sometimes involved with strangers and even lead to too much emotional connection and shared private information and sometimes lead to blackmailing in the form of money or sex etc.

IV. Some past cases of the misuse of social media.

- Abuse of social media where mob lynchings were reported in various parts of the country consequent to circulation of wild rumours on WhatsApp as a result of which innocent people have lost their lives
- A student from Bareilly was arrested by Rampur police for sharing an “objectionable” post on Facebook against a senior politician and state Urban Development Minister. A team of crime branch sleuths picked him up from his house and he had been remanded in judicial custody for 14 days.
 - An FIR has been lodged against an MP by Delhi police on the directions by a city court for his alleged “hate speech”. The FIR includes charges under Section 153 A (promoting enmity between groups), 504 (intentional insult with intent to provoke breach of peace) and 120B (criminal conspiracy) as well as sections 66A, 66F and 67 of the IT Act that relate to “objectionable” speech and cyber terrorism.
- An FIR has been lodged against controversial Bangladeshi writer for allegedly hurting religious sentiments (on Twitter) following a complaint in Bareilly district of Uttar Pradesh.
 - Two girls were arrested over their Facebook post questioning the shutdown in the Mumbai city for funeral procession of one of the biggest leader of Maharashtra , which also leads to the attack of a clinic owned by one of the girls uncle.
 - A political cartoonist based in Kanpur was charged with sedition and insulting national emblems in his cartoons. The Mumbai Police cyber wing had blocked the website of the cartoonist which leads to a widespread debate on freedom of expression in India.
 - A Kerala youth was arrested for posting abusive Facebook comments against the prime minister of India. The case was registered at Anchal police limits in Kollam district after he had posted three abusive posts on Facebook.
 - The Kerala Police cyber cell on arrested two youths on charges of circulating an e-mail, which contained a picture of a palatial mansion.
 - In 2012, multiple Twitter users threatened one lady Indian writer, poet and activist after she discussed something that hurts the sentiment of a larg group of people. The writer was threatened with acid attacks and televised gang rape.
 - In 2013, one lady Indian journalist shared some of her views on secularism and communalism in the Indian political system in her twitter handle. After that she was threatened with rape by Twitter users. The twitter users went to the extent of discovering her daughters school and name and written threatening messages.After that the lady journalist subsequently stopped sharing her personal views on Twitter.
 - A prominent Delhi-based women’s rights activist was harassed during a 2013 online chat about violence against women on news website Rediff by a person using the handle @RAPIST, until she exited the discussion.
 - Separately, in 2015, two Indian actresses criticised the prime minister’s social media initiative #SelfieWithDaughter. Both were sent abusive language and violent threats on Twitter.
 - A woman BuzzFeed’s India editor, was subject to rape threats after she praised Pakistani players on Twitter during a 2015 India-Pakistan cricket match. Many Indian

men sent participated in the abusive comments, including writers Chetan Bhagat and Suhel Seth.

- A lady journalist has been called India's most trolled woman. She was routinely harassed for her online comments, escalated on Twitter and Facebook in December 2015 after she described being sexually abused as a child in her book *This Unquiet Land*. Among other abusive terms, critics called her "antinational," a right-wing slur.

- In 2015, one journalist of Media One Group from the southern state of Kerala, published a personal account of child sexual abuse at a religious school in the southern city of Kozhikode on Facebook. Over 1,700 Facebook users shared her account, but it also attracted abuse from members of the one community, many of whom reported her Facebook account for violating community guidelines, with the result that it was temporarily blocked.

- In a series of incidents in 2015, Facebook users attacked one lady Indian social activist, first for criticising a leader of one of the political party, and then the India's late President APJ Abdul Kalam. Trolls attempted to hack her account, created a fake Facebook profile depicting her as a sex worker, and directed sexualized abuse at her children. Facebook temporarily suspended her profile after one of her abusers reported her for violating their real name guidelines, since she had withheld her last name, which indicates her caste.

- An Indian playback Singer found herself being attacked by online trolls after she criticized one of the superstars of Bollywood. "Women thrashed, people run over, wild life massacred and yet #hero of the nation. 'Unfair'. India full of such supporters," she tweeted.

- A US-based Indian woman took on her cyber blackmailer with a public Facebook post that went viral and eventually helped in nabbing the blackmailer.

And the number goes on from spreading mms to blackmailing and from sharing message and pictures to making derogatory comment etc.

V. Cyber crime.

All the criminal related online activities fall under the cyber crime. The crime may be small or large but crime means crime only. The crime can be bullying a friend to spreading fake news to tarnishing the image of some individuals to hacking or attacking the security system of a country to fraud and harassing someone. Cyber crime is an offensive act and there are enough law available to deal with any type of online crime. Injecting virus is also a crime. Identity thief is also a crime. Creating fake account is also a crime. Trying to extract money blackmailing to reveal private secret is also a crime. Harassing messaging and calling is also a crime. Promising job and asking money in return is also a crime. Whatever the online activities that are beyond the Indian IT act are all considered under the cyber crime.

VI. Laws against cybercrime.

There are enough numbers of act and regulations and law to curb and stop the misuse of social media. Even if there is enough law to protect the misuse of social media, it is very difficult to control everything because internet is totally virtual and it is very difficult to trace everything. The reason is anyone from anywhere can access to any server from across the world and also there is no any confirmation that the said person is he/she only and he/she is from india only.

Even though it is difficult to contain everything but the government of india is very much succesful till date to contain such misuse of social media. Below I am listing some of the necessary steps taken by the government of india to curb the social media and internet misuse.

VII. Legal Provisions to deal with objectionable content on social media platform:

Social media companies are expected to remove any unlawful content under Article 19(2) of the Constitution of India as and when brought to their knowledge either through a court order or through a notice by the government or its agency.

- The IT Act 2000

Provisions for blocking and removal of content

➤ Blocking of websites/information :

Section 69A of the Information Technology Act, 2000 provides for blocking of websites/webpages in specific conditions of –

- Defence of India
- Sovereignty and integrity of India
- Security of State
- Foreign relations with foreign states

Public Order for preventing incitement to the commission of any cognizable offense relating to above.

- Disabling or removal of information by intermediaries :

The intermediaries or the owner of the social media websites are required to disable or remove any unlawful information as and when brought to their knowledge through a court order or by through appropriate Government or its agency under provisions of section 79(3)(b) of IT Act, 2000. Various law enforcement agencies notify the social media sites for removal of contents / disabling of contents under provisions of section 79(3)(b) of IT Act, 2000.

The Information Technology (Intermediary Guidelines) Rules 2011 notified under section 79 of the IT Act requires that the Intermediaries shall observe due diligence while discharging their duties and shall inform the users of computer resources not to host, display, upload, modify, publish, transmit, update or share any information that is harmful, objectionable, affect minors and unlawful in any way.

Blocking Compliance of section 69A of IT Act, 2000

(i) This act allows the social media owners to block any account and any sites that is against the indian penal act. It can block the url.

The legal provisions that are currently available with regard to the conduct of intermediaries, which includes various social media platforms, in order to respond to such developments are in the IT Act, 2000. This is proposed to be done by strengthening the implementation aspects section 79 of IT Act, 2000.

The government of india is always in accordance with the constitution and whatever the provision that are under the constitution are all followed accordingly. The government of india is committed to freedom of speech and expression and privacy of its citizens as mentioned in the Constitution of India. Government does not regulate content appearing on social network platform.

These social network platforms are required to follow due diligence as provided in section 79 of the Information Technology Act 2000 and the Rules notified therein. They have also to follow Article 19(2) of the Constitution and ensure that their platforms are not used to commit and provoke terrorism, extremism, violence and crime.

The ministry takes down such information from the internet when it is in violation of Section 69A of the Information Technology Act. The Section 69A of the Information Technology Act empowers the government to block any information generated, transmitted, received, stored or hosted in any computer resource in the interest of the sovereignty and integrity of India, defence of India, security of the state, friendly relations with other countries or public order or for preventing incitement to the commission of any cognizable offence relating to such matters.

The increase in the numbers of social media uses, there are increasing number of incidents of reporting of content having potential to incite violence and hatred. Government has a well-defined system of reporting and blocking for public access within the framework of section 69A of the Information Technology (IT) Act, 2000 and the Rules there under,”

Social media companies are considered intermediaries under the IT Act, 2000. They are required to “observe due diligence while discharging their duties and have to inform the users of computer resources not to host, display, upload, modify, publish, transmit, update or share any information that is grossly harmful, harassing, defamatory, libelous, invasive of another’s privacy, hateful, or racially, ethnically objectionable, disparaging, or otherwise unlawful in any manner”.

They are also expected to remove any unlawful content under Article 19(2) of the Constitution of India as and when brought to their knowledge either through a court order or through a notice by the government or its agency.

- The consequences of the misuse of social media

There are two sides, one is the victim and the other is the convict. The loss is for both the parties. The lost can be lifelong or short term. The lost can be physical properties or long time mental agony. The lost can be to one individual or to a group or even may the loss of security of a country or may be the economical loss of a country. Sometimes the loss may be one's life also. The consequence of the misuse of social media is to fall into legal trouble and sometimes jail. So no one should use the negative advantage of social media. It is not like the user can carry out certain criminal activities and move away freely. The hand of the law is very long and can reach any computer at any place. And there are many national and international law to tackle online crimes and so it is not easy for criminal to escape uncaught. And not only going to jail, the criminals name will be totally affected in the society and their reputation and image will be totally spoiled. Everything is not as easy as it seems to be. So no user should misuse the social media because there will be huge loss in multiple possible ways.

VIII. Few tips which can prevent misuse of social media.

The popularity of social media has been soaring day by day and there are tremendous positive points as well as negative points. Due to its virtual nature and due to its limitless accessibility, many people have started misusing it either for fun or to tarnish the image of individual or to target the security and economy of a country. Here I am listing some of the idea to prevent misuse of social media.

The most important group is the people whose age are below 18. Parents have to give adequate knowledge to their children the proper use of social media and what to share and what not to and which particular links should be opened and which not. Parents also should give proper education to their children about the consequences of doing what is not permitted and also should guide their children to use only for good and productive things and should advice their children to not use longer and should give some particular time to use the social media. Parents should check

the online activities of their children and what they are sharing and whether their children are involving in any online bullying and irrelevant activities etc.

Parent should advise their children to not trust the pictures use as profile pictures and parents should also advise their children to not accept friends request coming from unknown people and parents should also advise their children to not reveal any personal or private information online and to not share unnecessary things.

Parents should teach their children what the meaning of social media is and when and for what purpose we can use the social media and whom to be trusted and whom not. And parents should also teach their children the consequences of sharing personal details and pictures or videos etc. And parents should also teach their children to not fall on the online fraudsters hand.

And the social media service providers should compulsorily use an audio presentation which contains all the necessary precautions to take when using social media so that users can listen to that and know all the rules and regulations to be followed when using the social media.

With or without the knowledge of the children, the parents should continuously monitor the online activities of their children so that they can turn their children to the right track always.

Parents should teach their children the chances of misinterpreting and misusing the different things their children shared on the social media. And parents should teach their children that once something is shared, that exist permanently and so to comment or share something thinking twice.

To not pressurize the children for using social media, parents can form a group where their children are there and so like this their children can enjoy the social media and also use it safely and nicely.

Parents should advise children the health affects of using computer or mobile for longer period of time and also lose in their study etc.

The next group of people is the adult. Adult includes all types of employees and all types of businessman and celebrities and politician etc. Their should always be a separate session when anyone first sign up in any social media. The session should be either in the form of video or audio and should contain everything related to the use of social media from positive to negative and the chances of falling on the legal issues and loss of personal or government property and the chances of affecting a country etc. Like this users can be awared about the use of social media.

The admin of the social media service provider, the system admin of the particular organization or the admin who is running a cyber café should always monitor the activities of the user and the admin should properly record the internet use activities of the user. This will tremendously decrease the misuse of socil media.

The social media tool kit can be utilized by an organization that brings together various policies and tools. It should be mandated to sign up to any websites by using the personal identity card containing valid serial number and clear picture and address. It will tremendously reduce the misuse of any service available on the internet. Because before making any mistake, the user will think multiple times the consequences. And at the same time one subject should be there in the school and college syllabus which teach only about the ethics and moral related to the use of internet and what are the consequences of misuse of internet contents and the different law available to tackle the problems of internet misuse and also what are the different facilities available when anyone is in the middle of the cyber attack etc and which act to follow and whom to be approached etc. So like this also we can control the cyber crime and the misuse of internet.

IX. Manually tracing the cyber crime.

The tracing of cyber crime using manual technique is time consuming but very effective. First the complainant should give details of the issue to the cyber crime explaining everything from first to last clearly step by step. Then the cyber crime officers will study the case appropriately by collecting all the details evidence. Then the officers will give the priority to the case according to severity. Then the officers will start investigating to the root and find the people involved in the crime and then will produce in the court and the court will decide with the suitable punishment as according to the cyber crime act. Manual tracing of cyber criminal is very time consuming because the officers need to collect evidence from the electronics device one by one and also need to visit the possible crime location and sometimes criminal use proxy ip address and fake identity. But it is not like impossible, the cyber crime investigation agent will somehow come out with the truth. The authorized party can then find out the content and delete.

X. Automatic detection of obscene contents in social media.

I propose two approaches to handle the issue of any online abuse detection in India:

❖ Text mining based approach:

Text mining means nothing but extracting valuable information from many unstructured data available online so that we can divide those information according to its types and make useful in different applications. Text mining is also know as text analytics or text analysis. Text mining is one of the artificial intelligence technology. Text mining uses natural language processing to transform the unstructured data into structured data and this structured data is very useful and applicable in many applications directly without any further processing. And we can use this extracted structured data as an input to the machine learning algorithms. Text mining identifies and groups different information so that it becomes easier for the researcher and machine learning engineers to use it without any more transformation.

Among the many technology used by text mining, the most common one is the natural language processing. The structured data created by text mining can be integrated into databases, data warehouses or business intelligence dashboards and used for descriptive, prescriptive or predictive analytics.

Text mining is based on a variety of advance techniques stemming from statistics, machine learning and linguistics. Text mining uses different types of techniques to find the patterns among the unstructured data. The main goal of text mining is to find useful data so that any user can use those data without further extraction in finding some other information.

Text mining has a large number of uses to include text clustering, concept extraction, sentiment analysis, and summarization. Text mining uses natural language processing for easy and automatic extraction of useful and structured data from the unstructured data. We know that natural language processing is a part of artificial intelligence and so it is faster and easier for text miner to use natural language processing for converting unstructured data into structured data. As we know that natural language processing techniques are being used to make the text or any online information understood by the machine easily and so less human efforts are required.

Text mining helps organizations to make their employee focus on other important task because the mining of valuable information will be taken care by the machine learning algorithms and artificial intelligence technologies. Text mining algorithms make users to be able to extract information of their interests. May be negative text, positive text, vulgar text and everything the user needs.

When text mining and machine learning are combined, automated text analysis becomes possible. And we know that to make the machine intelligent we need to train the machine by feeding enormous numbers of data till the accuracy of prediction of the machine has become quite good and acceptable. Text mining combines with machine learning, statistics and other linguistic mechanism to create machine learning model.

❖ Text Mining Methods and Techniques:

There are different methods and techniques for text mining.

- Word frequency

Word frequency means finding the most frequently occurring text from the unstructured data set. From this we can easily find the vulgar and onscene contents.

I am not mentioning any of the vulgar words here.

- Collocation

Collocation can be bigrams or trigrams etc. Collocation means finding the words that occurred together. Bigrams means two words occurring together and trigram means three words occurring together. So using this technique also we can find objectionable words and sentences and objectionable words that give bad meaning when combined together.

- Concordance

Concordance is a technique to find the meaning of a sentence from a sentence. We know that there are certain good words but the context of the word changes with respect to the way the speaker speaks. So like this also we can find vulgar text.

- Text Classification

Here we are doing the text classification by using the natural language processing so that less manual error occurs and speed up the process of classifying the text and assign any new text correctly to the appropriate group.

Text classification is quite useful in finding the text that we want and we can group them together and now the classified text can be in research and academic and other business and scientific applications.

Text classification are used in the following frequently:

Topic analysis, Sentiment analysis, Language detection, and Intent detection.

- Topic Analysis: This technique is used to know the main theme behind a users text. We are analyzing the user thinking and what the user want to express by using this technique.
- Sentiment Analysis: this is one of the application of natural language processing and this technique is used to know the sentiment of the user like what the user is thinking and what the user wants to express and what is the intention of the user from the users text.
- Language Detection: It means detecting a text and knowing the language from that text. It is also very useful in knowing the meaning of the text. And this technique can be used in identifying the meaning of text also.
- Intent Detection: This is the technique to know the intention of the user from the text the user share. From the information the user shared, we can know the real intention of the user and what the user actually wants to say.

- **Text Extraction:** It is one of the text analysis technique. This technique extracts whatever the particular text we want by using a sentence or a keyword etc. Now by using this technique we can find the text that we want.

The following are some of the application of text extraction:

keyword extraction, named entity recognition and feature extraction.

- **Keyword Extraction:** We are using particular words to find the text we want. We can use keyword to group related text and this is quite useful to find particular and its related terms.
- **Named Entity Recognition:** It is useful in identifying particular person, company, organization, industries, and anything that is related to name.
- **Feature Extraction:** this technique is used to extract feature of any product, websites, brand, school, colleges, institutes, industries etc.

The combined techniques of text mining, machine learning and artificial intelligence is very useful in finding any text of demand and it is also very useful in finding patterns to easily find the information of our interest. It is efficient, fast and can do repeatedly again and again the accuracy increases with time and experience.

In text mining, the most important task is to gather data from different sources. The source may be news channel, blogs, social media, emails, chats, surveys, databases, social media, institutes, medicines, engineering, electronics products etc.

The next most important task of text mining is preparing data to give as input and to train a machine learning model. Text mining uses many natural language processing techniques while extracting data like:

Tokenization, normalization, stemming, lemmatization, removing stop words and removing noise.

- **Tokenization:**

It is the process of breaking down big paragraphs into sentences and sentences into words and words into characters.

Normalization:

To achieve normalization, we need to perform multiple task simultaneously. It involves converting lowercase characters to uppercase characters and vice versa, elimination of punctuations and writing the numbers using alphabets.

- **Stemming:**

It is used to change the different tenses of words into its base words. It leads to reduction in unnecessary calculation of the same words due to the different form of tenses.

- **Lemmatization:**

It is the process of different words into single words. This analyzes the word morphology and eliminates the ending of the words with different grammatical rules.

- **Removing stop words:**

It is nothing but the removal of words which does not give any specific meaning to any sentence other than making the sentence grammatically correct. So it is very important to remove such words to reduce the unnecessary calculation.

- Noise removal:

It means removal of words that are incomplete and that are not related to our interest topics. It also removes words or sentence that give no meaning. It gives more meaningful text to the machine.

XI. Machine Learning Algorithms that can be used for Text and Image Detection:

1. Logistic Regression.
2. Random Forest
3. SVM(Support Vector Machine)
4. Decision Tree
5. Naïve Bayes

XII. The outcome of this paper:

It is difficult to exactly extract all the obscene words and sentences and we will need the machine learning model to train more and more and with more experience, the accuracy of the model increases and it is very useful in making the internet clean and safe for all. It will lead to less crime and also less numbers of personal and private problems.

XIII. CONCLUSION

This paper is about the issues that many of the users are facing nowadays due to the misuse of the social media by many users. The misuse can be many things like injecting viruses, misleading in the name of profits, spreading of rumours and spreading hatred audio, video, text etc and using abusive and vulgar words that are against the norms of using social media and also using many such unacceptable words which cant be used in a platform like social media. The name social media itself says that social media is for all and just like a community group where there are family, friends. Relatives and near and dear ones. And this paper will help in eliminating vulgar contents and in blocking websites and users that don't follow the basic norms of social media.

XIV. REFERENCES

1. <https://link.springer.com/article/10.1007/s41060-017-0088-4>
2. <https://towardsdatascience.com/nudity-detection-and-abusive-content-classifiers-research-and-use-cases-d8886bf624e8>
3. https://faculty.ist.psu.edu/xu/papers/Chen_etal_SocialCom_2012.pdf
4. https://www.researchgate.net/publication/343079524_Detection_of_Offensive_Language_in_Social_Media_Posts
5. https://www.researchgate.net/publication/261394382_Detecting_Offensive_Language_in_Social_Media_to_Protect_Adolescent_Online_Safety
6. <https://www.frontiersin.org/articles/10.3389/fdata.2019.00008/full>
7. <https://www.aclweb.org/anthology/W17-3008.pdf>
8. <http://ceur-ws.org/Vol-2633/paper9.pdf>
9. <https://towardsdatascience.com/machine-learning-nlp-text-classification-using-scikit-learn-python-and-nltk-c52b92a7c73a>