**ABSTRACT**

This project offers a comprehensive understanding of Cyberbullying incidents and their corresponding offences combining a series of approaches reported in relevant Work. The implementation provides the opportunity to systematically combine various element or Cyberbullying characteristics. Additionally, a comprehensive list of Cyberbullying-related offences is put forward. The offences are ordered in a Deep Neural Network classification system based on specific criteria to assist in better classification and correlation of their respective incidents. This enables a thorough understanding of the repeating and underlying criminal activities.

**CHAPTER-1**

This chapter briefs about cyberbullying, its harmful effects and impact on society. It also discusses its association with the proliferation of the Web and growing usage of social media. It talks about the challenges of the chosen research area and the need of deep learning for cyberbullying detection in social media. This chapter also briefly introduces the fundamental concepts related to the research area. It provides a brief description of the key terminologies namely, social media, cyberbullying, and deep learning, followed by a summary of the chapter.

# Introduction

Social media has reshaped communication by facilitating healthy discussions and candid conversations in which people engage on the community-centric platform by sharing ideas, thoughts and information. As one of the most popular and modern means of communication, social networking sites provide a constructive platform for market research, decision-making process and government intelligence [1, 2]. Undoubtedly, its mass adoption, effortless availability and popularity can get users united in a very short time and allow gathering opinions from different people on an issue in just a click. But this virtual social world can also fuel and witness different anti-social activities such as scams, fake news, rumours and cyberbullying.

Cyberbullying (CB) is a form of manipulation, belittlement, and targeted abuse using mean-spirited messages and negative electronic postings [3]. It is the use of information technology networks by individuals to humiliate, mock, embarrass, insult, defame and criticize a target without any one to one contact. Cyberbullying can be as straightforward as sending mean, hurtful, rude texts or instant messages as devious as spreading secrets or rumours about people online. Though bullying in electronic form can have multiple-dimensions, such as exclusion, harassment, outing, trickery, cyber stalking, dissing, fraping, masquerading, trolling and flaming [4, 5], the obvious intention to hurt and harm is common. This inveterate nuisance creates mental, emotional and physical risks for the bullied. The targets (victims of cyberbullying) feel overwhelmed, powerless, vulnerable, unsafe, worthless, humiliated, isolated, depressed, embarrassed, vengeful and at times suicidal.

Technology (Web 2.0) allows the bullies to be anonymous, hard to trace and insulated from confrontation. To the targets of cyberbullying, it feels invasive and never- ending. An accurate detection can facilitate timely intervention by alarming the moderators to take countermeasures. But content moderation practices on these platforms by human moderators is often inconsistent and done in a non-transparent

manner. It also suffers from biasing and may apprehend freedom of expression online. Moreover, spotting bullying instances explicitly as well as spotting victims is tricky too. Blocking and reporting might augment the reality if the bully is within the same professional or personal community, for example, a classmate. Sadly, the scale and impact of cyberbullying can be seen across social media platforms even though its awareness is at an all-time high. Simultaneously, with huge volume and variety of user-generated content on complex social media platforms, the challenges to detect cyberbullying in real- time have amplified. The influx of content makes it challenging to timely regulate online expression. Moreover, the anonymity and context-independence of expressions in online posts can be ambiguous or misleading. Recently, as memes, online videos and other image-based, inter-textual content have become customary in social feeds; typo-graphic and info-graphic visual content has also become a considerable element of social data. Thus, cyber bullying, through varied content modalities is very common. At the same time, cultural diversities, country-specific trending topics hash-tags in social media, the unconventional use of typographical resources such as capitals, punctuation and emojis and easy availability of native language keyboards add to the variety and volume of user- generated content compounding the linguistic challenges in detecting online bullying posts.

Researchers worldwide have been trying to develop new ways to detect cyber bullying, manage it and reduce its prevalence in social media. Advanced analytical methods and computational models for efficient processing, analysis and modelling for detecting such bitter, taunting, abusive or negative content in images, memes or text messages are imperative. The automated cyberbullying detection has attracted growing interest over the past decade as it facilitates combating toxic online behaviour. A lot of research has been done on detecting cyberbullying in textual data using a myriad of features [5]. Many datasets have been made open-source to facilitate research enthusiasts.

As a classical problem in natural language processing (NLP), cyberbullying detection in real-time user generated content needs high-level semantic analysis. Most of the earlier attempts on cyberbullying detection rely on manual feature extraction methods [6]. Such methods are not only time-consuming and cumbersome, but often fail to correctly capture the meaning of the sentence. Few lexicon-based methods by maintaining a list of offensives, abusive and hateful words have also been used, but are quite limited in scope [7]. Recent research focuses on the application of deep learning models for various NLP tasks and has reported state-of-the-art results [8]. Basically, deep architectures are neural networks with multiple processing layers of neuron with each layer having a specific task [9]. Utilizing deep learning models trivializes the need of explicit feature extraction techniques as these models are highly skilful and fast in retrieval of essential features and patterns by themselves. With minimal human intervention these models report superior results than the conventional machine learning models. Various deep learning architectures have contributed significantly in

analytics of text [9]. Pre-trained word embedding’s like Word2Vec, GloVe, ELMo, fastText that represent text in vector forms and deep neural networks such as CNN, RNN, GRU, LSTM, CapsNet & hierarchical networks that automate the task of feature extraction demonstrate best practices for solving text classification problems [10, 11]. Deep architectures with better representation learning capabilities have been substantiating its relevance in this field with improved results. Assessing the user- generated content in social media could be rewarding for automatic cyberbullying detection using deep neural architectures.

Thus, as an effort to deal with the antagonistic online delinquency referred to as cyberbullying, this research computationally analysed the content, modality and language-use in social media using deep learning models. The research demonstrates the feasibility, scope and relevance of using deep learning models for cyberbullying detection in social media portals.

# Statement of Research Question and Research Objectives

The conventional methods used to analyse data are inadequate as unlike the traditional data, social media data is mainly unstructured and comprises multilingual text and in varied modalities such as audio, video, images, GIFs, Emojis’ etc. Moreover, the linguistic complexities of user-generated content in social media makes it even more intricate to tap and analyse information using contemporary tools. Novel approaches to information discovery and decision making which use multiple intelligent technologies such as machine learning, deep learning, artificial intelligence, natural language processing and image recognition among others are required to understand data & then generate insights.

**Statement of Research Question:**

***"Can the linguistic complexities of user-generated content in social media be computationally analysed using deep learning models for automated cyberbullying detection?”***

Based on the statement of research question, the following research objectives (RO’s) are

identified:

**Research Objective I**– *To detect cyberbullying in textual social media content using deep learning model.*

**Research Objective II**– *To apply deep learning model for cyberbullying detection in content modalities other than text i.e., multimodal social media content*.

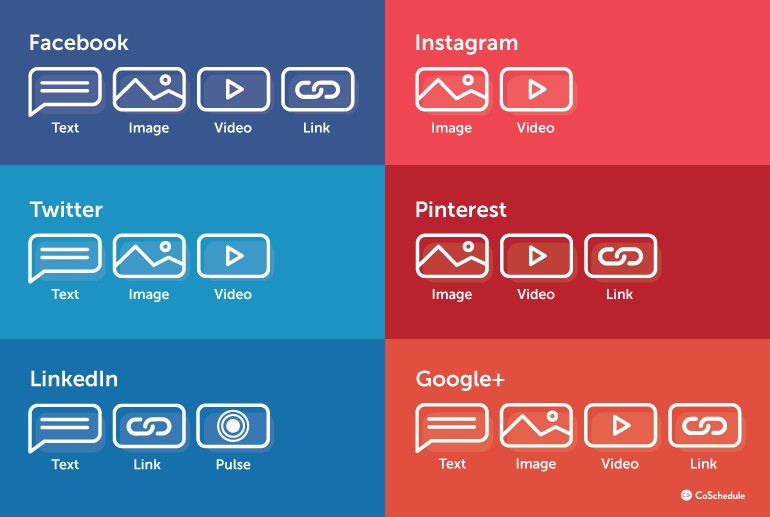
**Research Objective III**– *To computationally identify cyberbullying in mash-up social media content using deep learning model*.

# Social Media

Information is power, but without a means to distribute information, people cannot harness this power. Social media come up as a key player that gives a platform for expression and content distribution in today’s world [12]. The basic purpose of social media sites is to build interest, professional and interconnection-based virtual groups empowering better connections with other people all over the world. With the rapid growth of these sites (Instagram, Twitter, Snapchat, Tumblr, YouTube, Google+, Facebook, etc.), the netizen can share all type of social media data viz. text, audio, image, video utilizing the power of Internet without having ample information regarding the network topology and client-server architecture of Web. The social networking sites have given ‘everyone a voice’ but at the same time, we’re drowning in abundance, complexity of choices and unfortunately, the misappropriation or misdirection of influence. Moreover, when lots of individuals come together and that too from different countries, communities, races, ethnicities, gender, and varied age-groups, there are bound to be conflicts, controversies, and intimidation vulnerabilities. But this virtual social world can also fuel and witness different anti-social activities such as scams, fake news, rumours and cyberbullying [13]. That is, although social networking sites proffer numerous benefits as these facilitate participation and collaboration but on the flip side hate speech, social distrust, cyberbullying, identity theft, cyber-stalking and cascading of rumours and fake stories are some antithetical concerns associated with it. The pervasive reach of these sites has irrefutably triggered, contributed and exacerbated bullying.

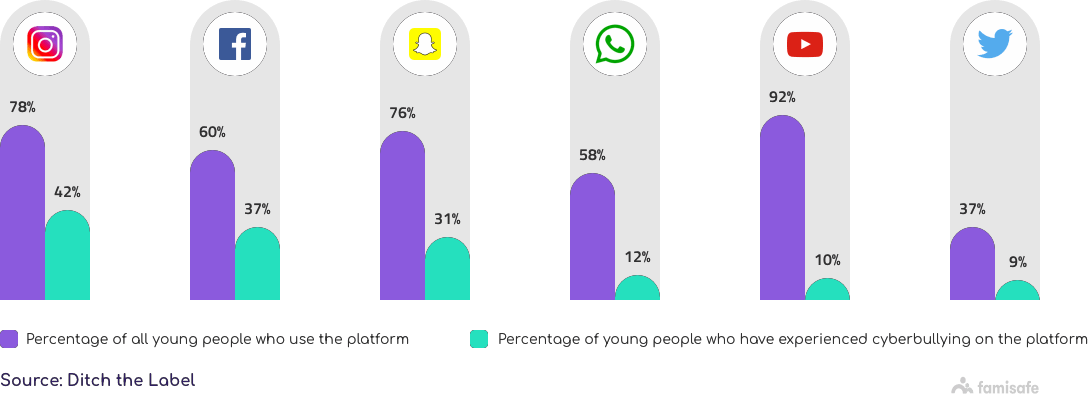
Social media may seem positive and safe, but it affects our daily lives more than we can think of. According to a study by Harvard University, “self-disclosure on social networking sites lights up the same part of the brain that also ignites when taking an addictive substance. The reward area in the brain and its chemical messenger pathways affect decisions and sensations” [14]. The overuse of social media can disrupt psychologically leading to social withdrawal, depression, anxiety and insomnia. Further, social media hacks and oversharing makes one’s identity extremely vulnerable.

Social media is inherently an informal way of communication with all kinds of multimedia content. The following figure 1.1 [15] depicts the multimedia types supported by popular social networking sites.



**Fig. 1.1.** Multimedia support by popular social networking sites

Social media dynamics keep changing with respect to increasing user base and user-activity which makes it a high dimensional, complex and ambiguous data space for analytical processing. Pertinent studies indicate that social media is one of the most favoured mediums by bullies and various factors such as socio-demography, physiological distress and time frames are related to cyberbullying. The massive volumes of human-centric, real-time, multimodal, heterogeneous and unstructured social media data makes manual detection intractable. Moreover, the social web applications/services are not restricted to the text-based data but extend to the partially unknown complex structures of image, audio and video. This fosters the need to develop intelligent tools and techniques for identifying, detecting and assessing cyberbullying from the available social media data to lower down its hazardous impact. Design and development of contemporary tools which tap and analyse online detrimental behaviour automatically from the high-dimensional social media are imperative. The substantial growth in the dimensionality, heterogeneity, subjectivity and multimodality of social media and the pressing need to timely curtail the damage instigated through cyberbullying, has fostered the need to devise automated mechanisms which detect such unfavourable activities. Social media has made cyberbullying a lot easier than it used to be due to it being much reckless in reach and virality that too with anonymity and without any restrictions. Social media cyberbullying is most prevalent in Instagram (42%), followed by Facebook (37%) and Snapchat (31%)1. Cyberbullying can be as straightforward as sending mean, hurtful, rude texts or instant messages as devious as spreading secrets or rumours about people online. Though bullying in electronic form can have multiple-dimensions, such as exclusion, harassment, outing, trickery, cyberstalking, dissing, fraping, masquerading, trolling and flaming [16, 17], the obvious intention to hurt and harm is common. This inveterate nuisance creates mental, emotional and physical risks for the bullied. The targets (victims of cyberbullying) feel overwhelmed, powerless, vulnerable, unsafe, worthless, humiliated, isolated, depressed, embarrassed, vengeful and at times suicidal. An accurate detection can facilitate timely intervention by alarming the moderators to take countermeasures. But content moderation practices on these platforms by human moderators is often inconsistent and done in a non-transparent manner. It also suffers from biasing and may apprehend freedom of expression online. Moreover, spotting bullying instances explicitly as well as spotting victims is tricky too. Blocking and reporting might augment the reality if the bully is within the same professional or personal community, for example, a classmate. Sadly, the scale and impact of cyberbullying can be seen across social media platforms even though its awareness is at an all-time high. Figure 1.2 shows how social media platforms are hotbeds of cyberbullying activities for young people.

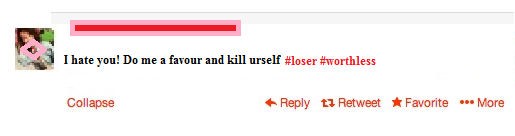


**Fig.1.2.** Statistics on share of social media platforms where cyberbullying occurs2

Typically, online bullying involves sending or posting harmful content or negative comments about a person. It intends to embarrass or humiliate a person in order to ruin his/her dignity, confidence and self-esteem [18]. The results of cyberbullying are dangerous and may affect the victim socially, mentally or psychologically. Hence, it is important to promptly detect cyberbullying in order to prevent it from becoming a global epidemic. Cyberbullying

Cyberbullying is defined as bullying an individual or a group of individuals using the Internet, mobiles or any other electronic device by sending inappropriate textual or non- textual multimedia messages in order to hurt or cause embarrassment [19]. The one who bullies is called a ‘bully’ and the other is said to be ‘victim’. The term ‘Cyberbullying’ was coined by Canadian educator and anti-bullying activist Bill Belsey in the year 2003 [20]. It is the repeated exposure of the negative actions on the part of one or more individuals in order to inflict humiliation, harassment, discomfort or injury upon another through the use of electronic medium [19] like emails, chat rooms, instant messaging, cell phones or by posting videos, audios, images etc. Bullying has been a part of human civilization history which involves hurting someone either by humiliating or harassing in any form, involving mental, verbal or physical damage. When this assault takes place in cyberspace, it is referred to as cyberbullying/ cyber-harassment/ cyber-victimization [21].

According to a study, nearly 43% of the teenagers in the United States are victims of cyberbullying [22]. It is more persistent way of bullying an individual in front of the entire online community especially within the social setting which can eventually lead to psychological, mental and emotional breakdown for the victim inculcating the sense of low self-esteem, low self-confidence, anger, depression, stress, loneliness, sadness, health degradation etc. [23]. Many of such intense cases have tragically ended in self-injury or suicides, underlining the grave nature of this critical issue [24]. The following figure 1.3 presents an example of cyberbullying from Twitter.



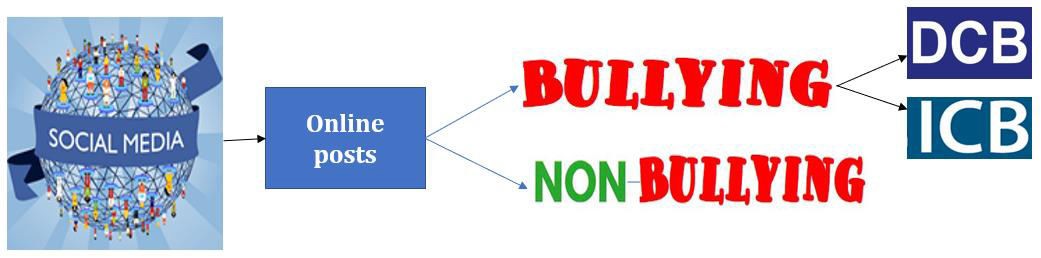
**Fig.1.3.** Example of cyberbullying

With technological advancement, the social freedoms that the networking sites give and larger audience, cyberbullying has spread manifolds affecting the individual not only limited to their workplace but also children and young adults in their daily lives. Anonymity further allows bullies to be more aggressive and offensive due to the reduced chance of being detected and punished, making it critical to efficiently detect cyberbullying behaviour in a real- time setting. This poses significant threat to the

physical and mental health of the victims making it a public health concern. Various studies have reported that victims of cyberbullying have lower self-esteem, higher levels of depression, suffer from behavioural issues and are addicted to substance abuse. Bullying victimization may trigger a sequence of events that results in suicidal behaviour. The first reported case of cyberbullying was of an American middle school student, Ryan Halligan of Vermont in 2003 [25]. Ryan was constantly bullied in person and online by his classmates and this bullying was attributed as his reason to commit suicide. As per the National Bullying Prevention Centre, ‘Every child on Facebook likely has a bullying story, whether as the victim, bully or as a witness’ [26].

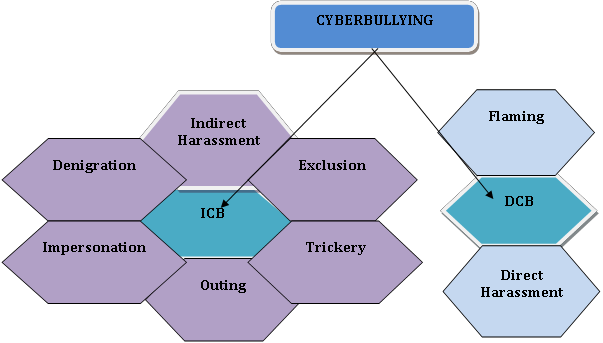
Netiquette refers to good manners on the Internet and treating other people on the Internet as you would like to be treated yourself. Unfortunately, some people use the Internet and/or mobile phones to offend or harass others. This is referred to as cyberbullying. Automated cyberbullying detection is a proactive strategic technology- based mechanism. It is a typical inherent classification problem of natural language processing where the intent is to classify the social media messages as either bullying or non-bullying. Cyberbullying is a multi-step process (for predictive analysis) comprising various sub-tasks such as data collection and it’s pre-processing; extracting and selecting relevant features and thereby classifying messages. The increasing use of social media at such a fast pace is adding both variety and volume to user-generated content, owing to which the manual classification (as either CB or non-CB) has become quite intractable. Simultaneously, it is generating an enormous number of features also. Choosing the most appropriate feature is a challenging task [27] and it influences the overall classification accuracy as well. This necessitates assessing & examining novel computational approaches that show better representation learning capabilities and simplify the feature selection process with enhanced classification accuracy and ensure result comprehensibility as well.

It is often characterized as a predictive learning model in the social setting which detects the presence of cyberbullying in an online post (textual/non-textual) so that it does not inflict seriously or damage the victim’s emotional, psychological and social state. The posts classified as bullying can further be divided into two categories, namely, direct cyberbullying (DCB) and indirect cyberbullying (ICB) [28], as shown in figure 1.4. DCB involves direct sending of harmful content to a person either via email or SMS etc. ICB comprises of posting harmful contents about any person on social media or sharing it with others, for example posting an improper photograph of someone on Facebook is an example of ICB.



**Fig.1.4.** Classification of social media posts

The direct and indirect cyberbullying is categorized into the different types as depicted in figure 1.5.



**Fig.1.5.** Types of cyberbullying

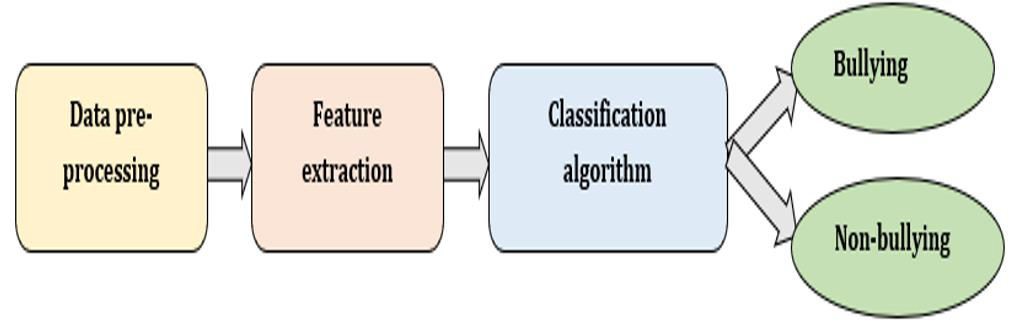
ICB messages are further categorized into six types [28]. These include Indirect Harassment, Denigration, Impersonation, Outing, Trickery and Exclusion, whereas DCB includes Flaming and Direct Harassment.

The following table 1.1 briefly explains these types of cyberbullying.

**Table 1.1.** Types of cyberbullying

|  |  |
| --- | --- |
| **Types of cyberbullying** | **Details** |
| **Flaming** | It is an online fight between individuals. They usually exchange vulgar electronic messages. People fighting on online forums exchanging obscene messages is an example of Flaming. |
| **Direct Harassment** | It is directly harassing a person either by insulting or threatening him or her via messages. It includes only two parties- the one who bullies and the other who is bullied. Threatening or harassing a person either by sending email or SMS directly is an example of direct harassment. |
| **Indirect Harassment** | It is indirectly harassing a person either by insulting or threatening him or her via messages posted online. It includes many parties. Posting embarrassing photos on social media in order to harass the other person indirectly is an example of indirect cyberbullying. |
| **Denigration** | It is spreading hearsay or rumours about others in order to ruin their reputation. It puts the status of the cyber-victim on stake. Posting skewed contents on forums or blogs etc. in order to turn down cyber- victim’s reputation is an example of denigration. |
| **Impersonation** | It is acting or pretending as another person and then doing anti-social activities in order to embarrass or damage his or her reputation. Imitating cyber-victim either by creating a fake profile or through hacking and sending messages that may instigate other users to attack the victim is an example of impersonation. |
| **Outing** | It is sharing private information of a cyber-victim without his or her consent in order to hurt the victim. Posting a humiliating picture of someone in order to hurt the cyber victim is an example of outing. |
| **Trickery** | It is obtaining sensitive information about a user by faking the trust of cyber victims and then eventually violating that trust. Obtaining a personal video by faking as a close friend and then posting it online is an example of trickery. |
| **Exclusion** | It involves the exclusion of the cyber victim from online communities or groups etc. Excluding a person knowingly from a WhatsApp group is an example of exclusion. |

The elusive nature of cyberbullying undermines the self-esteem of the cyber victim, affecting him or her mentally, socially and psychologically. Automated detection model consists of multiple tasks which identify and classify posts as bullying or not. Considered as a generic classification problem, a typical cyberbullying detection process extracts the features from the pre-processed data and classifies the posts accordingly as shown in the figure 1.6 below.



**Fig.1.6.** Generic cyberbullying detection process

The pre-processing phase includes cleaning the acquired data by removing unwanted URL’s or strings etc., handling missing values, correcting words etc. and then transforming it into a representation suitable for feature extraction. After pre-processing, features such as keywords depicting bad/nasty/rude/abusive/hateful/attacking words, N-grams, pronouns, skip-grams are extracted. Next phase uses supervised learning techniques to classify the messages as either containing bullying content or not.

Cyberbullying is primarily associated with the utilization of digital media in order to bully someone. It has grown to a level where it is seriously affecting and damaging individual’s lives where social media forums play a key role by providing a fecund means for bullies and the one’s using such portals are more vulnerable to attacks. Nowadays, the Internet has drastically reformed the way people express their views-opinions-thoughts on social media. People rely more on the use of social forums like Twitter, Facebook, Formspring.me, MySpace, Ask.fm etc. for sharing their views & opinions which results in producing unprecedented volume of user generated online data that is available generally in the form of tweets, blog posts, reviews, question-answering forums etc. The heavy dependence of the mass on such multimedia content for appropriate opinions shows the increasing relevance of the utilization of Web 2.0 technologies and tools in our daily lives. Hence, we can say that social media has global reach and has become widespread. Its pervasive reach has in return given some unpremeditated consequences as well where people are discovering illegal & unethical ways of using such communities. One of its most severe upshots is known as cyberbullying where individuals are searching new means to bully one another over the Internet. It has grown as a social menace that puts a negative effect on the minds of both the bully and victim. It is more persistent wayof bullying a person before an entire online community especially when we talk in terms of social networking websites which can ultimately results in psychological and emotional breakdown for the cyberbullying victim developing the feeling of low self- confidence, depression, stress, anger, sadness, health degradation, loneliness, suicides etc. All this has gradually increased the linguistic challenges associated with the ‘user- generated real time social media content’ which further encourages the need to search for enhanced classification methods and paradigms which can cater well with cyberbullying detection in social media.

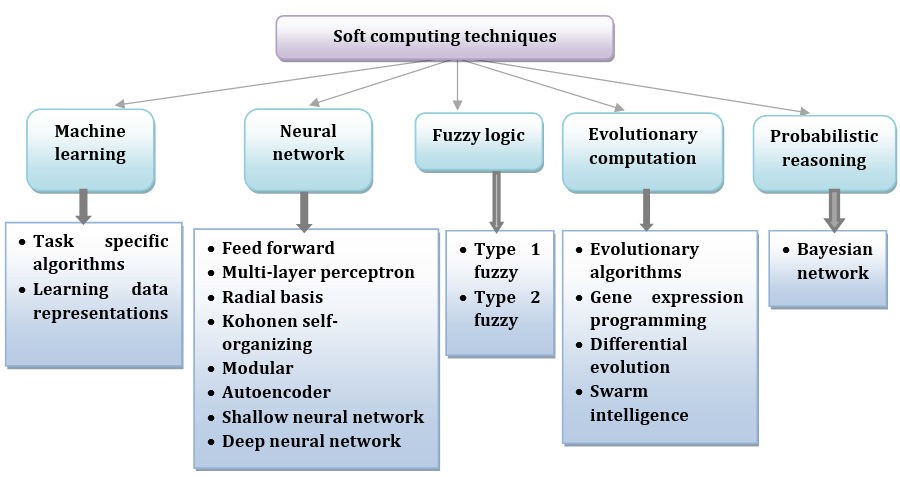
# Deep Learning

The volume and variety of user-generated content on complex social media platforms have amplified the challenges to detect cyberbullying in real-time. The influx of content makes it challenging to timely regulate online expression. Moreover, the anonymity and context-independence of expressions in online posts can be ambiguous or misleading. Recently, as memes, online videos and other image-based, inter-textual content have become normal in social feeds; typo-graphic and info-graphic visual content has also become a substantial element of user-generated data. Thus, cyber bullying, through varied content modalities is very common. At the same time, cultural diversities, country- specific trending topics hash-tags in social media, the unconventional use of typographical resources such as capitals, punctuation and emojis and easy availability of native language keyboards add to the variety and volume of user- generated content compounding the linguistic challenges in detecting online bullying posts. Researchers worldwide have been trying to develop new ways to detect cyber bullying, manage it and reduce its prevalence in social media. Advanced analytical methods and computational models for efficient processing, analysis and modelling for detecting such bitter, taunting, abusive or negative content in images, memes or text messages are imperative. The automated cyberbullying detection has attracted growing interest over the past decade as it facilitates combating toxic online behaviour. A lot of research has been done on detecting cyberbullying in textual data using a myriad of features [29]. Many datasets have been made open-source to facilitate research enthusiasts.

As a classical problem in natural language processing (NLP), cyberbullying detection in real-time user generated content needs high-level semantic analysis. Most of the earlier attempts on cyberbullying detection rely on manual feature extraction methods [30]. Such methods are not only time-consuming and cumbersome, but often fail to correctly capture the meaning of the sentence. Few lexicon-based methods by maintaining a list of offensives, abusive and hateful words have also been used, but are quite limited in scope [31]. Recent research focuses on the application of deep learning models for various NLP tasks and has reported state-of-the-art results [32]. Basically, deep architectures are neural networks with multiple processing layers of neurons with each layer having a specific task [33]. Utilizing deep learning models trivializes the need

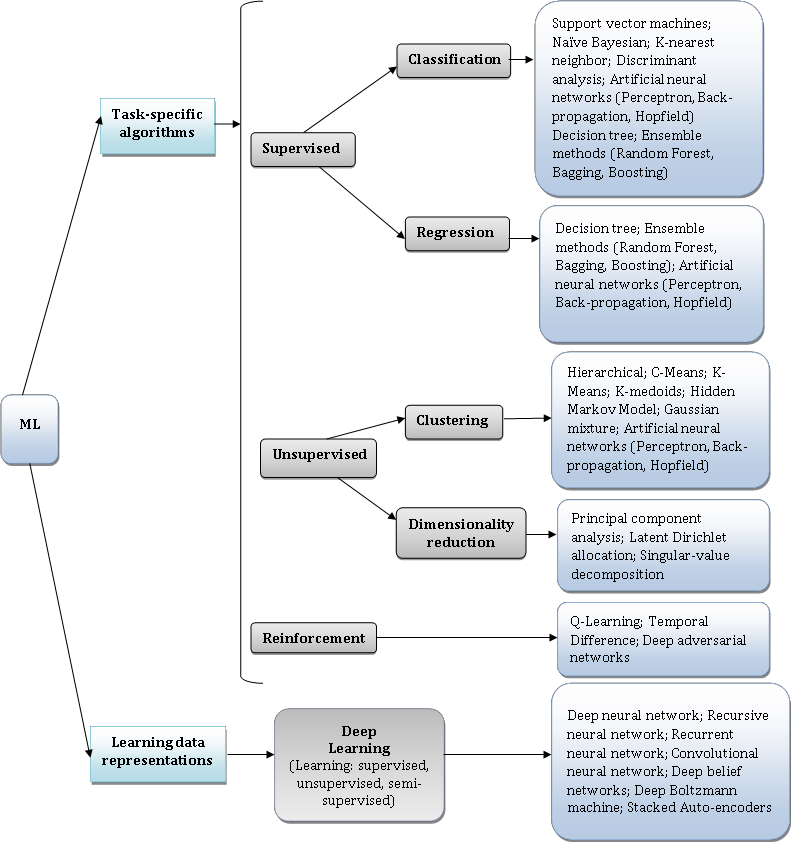
of explicit feature extraction techniques as these models are highly skilful and fast in retrieval of essential features and patterns by themselves. With minimal human intervention these models report superior results than the conventional machine learning (ML) models.

Deep learning (DL) is considered as a part of the broader family of ML based on learning data representations, in contrast to the task-specific algorithms and where learning can be supervised, semi-supervised or unsupervised. DL entails techniques such as deep neural network (DNN), recurrent NN, CNN, deep-belief networks etc., whereas NN is one of the sub-types of SC techniques [34] which includes feed forward; MLP; deep NN (DNN); radial-basis etc. (as shown in figure 1.7 and 1.8).



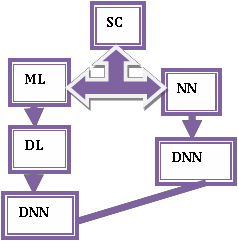
**Fig.1.7.** Categorization of soft computing techniques

From figure 1.7, it is apparent that soft computing is a ‘blanket term’ comprising several techniques which are themselves interrelated to one another. Also, referred to as computational intelligence techniques, soft computing (SC) techniques are categorized into machine learning (ML), neural networks (NN), evolutionary computation, fuzzy logic and probabilistic reasoning) [34].



**Fig.1.8.** Categorization of machine learning techniques

From figures 1.7 and 1.8, we can infer that DL is considered as a sub-part of ML. Thus, it can be inferred that SC, ML and DL are inter-connected to each other (as shown in figure 1.9).



**Fig.1.9.** Relation between SC, ML and DL

These deep architectures have better representation learning capabilities. They perform automatic feature extraction for the desired outcomes and are also fast. Assessing the user-generated content in social media could be rewarding for automatic cyberbullying detection using the deep neural architectures. This research demonstrates the viability, scope & significance of using DL models for CB detection in social media portals. Application of deep learning models for cyberbullying detection in social media is an upcoming area of research for finding, exploring and analysing the extensibility of human-based expressions.

**1.3 OBJECTIVE**

* To effectively develop a system to help social media platforms detect and identify abusive comments.
* To make use of Deep Neural Network, a regression machine learning algorithm for accurate detection of abusive comments.

**CHAPTER -2**

**LITERATURE SURVEY**

**2.1 INTRODUCTION**

The following shows survey on social media abuse detection. The most popular of the existing techniques used for power forecasting.

**1) A survey of data mining techniques for analyzing crime patterns**

**AUTHORS:**  U. Thongsatapornwatana

In recent years the data mining is data analyzing techniques that used to analyze crime data previously stored from various sources to find patterns and trends in crimes. In additional, it can be applied to increase efficiency in solving the crimes faster and also can be applied to automatically notify the crimes. However, there are many data mining techniques. In order to increase efficiency of crime detection, it is necessary to select the data mining techniques suitably. This paper reviews the literatures on various data mining applications, especially applications that applied to solve the crimes. Survey also throws light on research gaps and challenges of crime data mining. In additional to that, this paper provides insight about the data mining for finding the patterns and trends in crime to be used appropriately and to be a help for beginners in the research of crime data mining.

**2) Risk terrain modeling: Brokering criminological theory and GIS methods for crime forecasting**

**AUTHORS:** J. M. Caplan, L. W. Kennedy, and J. Miller

The research presented here has two key objectives. The first is to apply risk terrain modeling (RTM) to forecast the crime of shootings. The risk terrain maps that were produced from RTM use a range of contextual information relevant to the opportunity structure of shootings to estimate risks of future shootings as they are distributed throughout a geography. The second objective was to test the predictive power of the risk terrain maps over two six‐month time periods, and to compare them against the predictive ability of retrospective hot spot maps. Results suggest that risk terrains provide a statistically significant forecast of future shootings across a range of cut points and are substantially more accurate than retrospective hot spot mapping. In addition, risk terrain maps produce information that can be operationalized by police administrators easily and efficiently, such as for directing police patrols to coalesced high‐risk areas.

**3) Using geographically weighted regression to explore local crime patterns**

**AUTHORS:** M. Cahill and G. Mulligan

The present research examines a structural model of violent crime in Portland, Oregon, exploring spatial patterns of both crime and its covariates. Using standard structural measures drawn from an opportunity framework, the study provides results from a global ordinary least squares model, assumed to fit for all locations within the study area. Geographically weighted regression (GWR) is then introduced as an alternative to such traditional approaches to modeling crime. The GWR procedure estimates a local model, producing a set of mappable parameter estimates and t-values of significance that vary over space. Several structural measures are found to have relationships with crime that vary significantly with location. Results indicate that a mixed model— with both spatially varying and fixed parameters—may provide the most accurate model of crime. The present study demonstrates the utility of GWR for exploring local processes that drive crime levels and examining misspecification of a global model of urban violence.

**4) Language usage on Twitter predicts crime rates**

**AUTHORS:** A. Almehmadi, Z. Joudaki, and R. Jalali

Social networks 1 produce enormous quantity of data. Twitter, a microblogging network, consists of over 230 million active users posting over 500 million tweets every day. We propose to analyze public data from Twitter to predict crime rates. Crime rates have increased in the past recent years. Although crime stoppers are utilizing various technics to reduce crime rates, none of the previous approaches targeted utilizing the language usage (offensive vs. non-offensive) in Tweets as a source of information to predict crime rates. In this paper, we hypothesize that analyzing the language usage in tweets is a valid measure to predict crime rates in cities. Tweets were collected for a period of 3 months in the Houston and New York City by locking the collection by geographic longitude and latitude. Further, tweets regarding crime events in the two cities were collected for verification of the validity of the prediction algorithm. We utilized Support Vector Machine (SVM) classifier to create a model of prediction of crime rates based on tweets. Finally, we report the validity of prediction algorithm in predicting crime rates in cities.

**5) Self-organised critical hot spots of criminal activity**

**AUTHORS:** H. Berestycki and J.-P. Nadal

In this paper1 we introduce a family of models to describe the spatio-temporal dynamics of criminal activity. It is argued here that with a minimal set of mechanisms corresponding to elements that are basic in the study of crime, one can observe the formation of hot spots. By analysing the simplest versions of our model, we exhibit a self-organised critical state of illegal activities that we propose to call a warm spot or a tepid milieu2 depending on the context. It is characterised by a positive level of illegal or uncivil activity that maintains itself without exploding, in contrast with genuine hot spots where localised high level or peaks are being formed. Within our framework, we further investigate optimal policy issues under the constraint of limited resources in law enforcement and deterrence. We also introduce extensions of our model that take into account repeated victimisation effects, local and long range interactions, and briefly discuss some of the resulting effects such as hysteresis phenomena.

**CHAPTER - 3**

**SYSTEM ANALYSIS**

* 1. **EXISTING SYSTEM**

With the prevalence of the Internet, online reviews have become a valuable information resource for people. How- ever, the authenticity of online reviews remains a concern, and deceptive reviews have become one of the most urgent network security problems to be solved. Review spams will mislead users into making suboptimal choices and inflict their trust in online reviews. Most existing research manually extracted features and labeled training samples, which are usually complicated and time- consuming.

This paper focuses primarily on a neglected emerging domain - review, and develops a novel unsupervised spam detection model with an attention mechanism. By extracting the statistical features of reviews, it is revealed that users will express their sentiments on different aspects of movies in reviews. An attention mechanism is introduced in the review embedding, and the conditional generative adversarial network is exploited to learn users’ review style for different genres of comments. The experimental results demonstrate the superior performance of the proposed approach.

* 1. **DISADVANTAGES OF EXISTING SYSTEM**
* The existing system focuses on identifying offensive comments ,it is very slow and high false positives will be observed.
* It assumes that all predictors are independent. It’s estimations can be wrong in some cases so we cannot take probability outputs seriously.
* The system is not very accurate.
* Naive Bayes assumes that all predictors (or features) are independent, rarely happening in real life. ...
* This algorithm faces the 'zero-frequency problem' where it assigns zero probability to a categorical variable whose category in the test data set wasn't available in the training dataset.

**3.3 PROPOSED SYSTEM**

User Base for social media such as, twitter has increased multiple times in recent past. Also there is high increase in unstructured data in the means of reviews and comments in online portal. With these huge data growing there is lot of scope to use big data technologies to analyse these data.In our project, we propose to handle unstructured data using machine learning algorithms. We use these technologies to address cyberbullying in Twitter. In the proposed system, Deep Neural Network, a machine learning algorithm is used to train the dataset consisting of a collection of abusive comments.By accurately predicting the abusive comments used the system can be used by social media platforms to block such abusive comments.Thus, improving the user experience for using and accessing social media platforms. This project aims to contribute toward better understanding Cyberbullying by proposing a schema-based Cyberbullying incident description that: identifies the feature s of a Cyberbullying incident and their potential elements and provides a DNN Based offence classification system based on specific criteria.The proposed schema can be extended with a list of recommended actions, corresponding measures and effective policies that counteract the offence type and subsequently the particular incident. This matching will enable better monitoring, handling, and moderating the various Cyberbullying offences and their incarnation in the form of specific incidents.

**3.4 ADVANTAGES OF PROPOSED SYSTEM:**

* A technological solution for detecting abusive comments on social media and low false positives will be observed.
* Deep Neural Network is used which provides accurate detection.

**CHAPTER-4**

**SYSTEM DESIGN**

**DETAILED DESIGN OF THE PROJECT:**

This chapter describes the overall and the detailed architectural design. It also describes each module that is to be implemented along with Data Flow diagram.

**4.1 ARCHITECTURE DIAGRAM:**

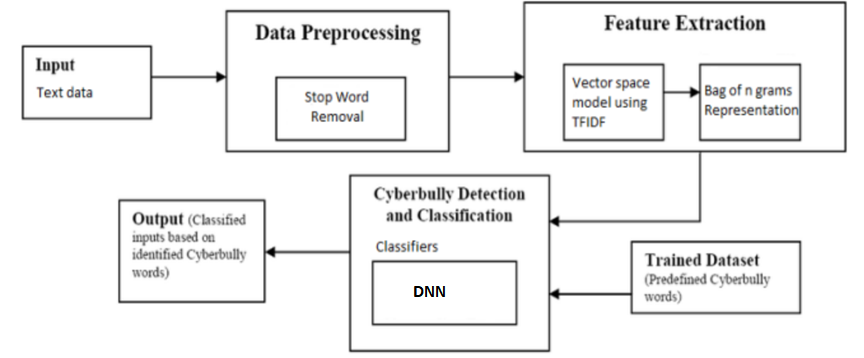


Figure 4.1 System Architecture

**4.2 WORKING:**

The aim of this project is to investigate and implement algorithms that could possibly detect and identify the abusive comments on social media . Data mining techniques and machine learning algorithms can be used for the prediction and detection of abusive comments.In this project the initial step will be the collection of different datasets from internet on various abusive comments perpetrated over social media internet which will be helpful in analysing the abusive comments then those dataset will be aligned accordingly. Then it will undergo a process called separation of datasets into training as well as testing where the training datasets will be used to train the model as well as testing will be used for evaluating the model. Then dataset pre-processing will be done which will align all the datasets into a specific category. There exist several regression algorithms in machine learning to develop an abusive comment detection model such as Deep Neural Network algorithm . In the proposed system, Deep Neural Network, a regression machine learning algorithm is used to train the dataset consisting of data on the abusive comment detection and identification.By accurately detecting and marking abusive comments over social media it can be very helpful in improving the social media experience.

**4.3 MODULE DESCRIPTION:**

1. Dataset Collection Module

Separation of Dataset

Labelling the Dataset Prior Training

1. Splitting of Dataset

80% Training

20% Testing

1. Dataset Pre-Processing Module
2. Training with Random Forest Algorithm

Validation and Evaluation

Merging of data

1. Comment Detection

Abusive Comment

Non-Abusive Comment

**4.3.1 DATASET COLLECTION MODULE:**

A data set is a collection of data. Machine learning has become the go-to method for solving many challenging real-world problems. It’s definitely by far the best performing method for prediction tasks. These machine learning machines that have been working so well need fuel lots of fuel; that fuel is data. The more labelleddata available, the better our model performs. The idea of more data leading to better performance has even been explored at a large-scale by Google with a dataset of 300 Million images! When deploying a machine learning model in a real-world application**,** data must be constantly fed to continue improving its performance. And, in the machine learning era, data is very well arguably the most valuable resource. There are three steps of collecting data.

**Classification**. When an algorithm to answer binary yes-or-no questions or to make a multi-class classification (grass, trees, or bushes; cats, dogs, or birds etc.)

**Regression**. For an algorithm to yield some numeric value. For example, if you spend too much time coming up with the right price for your product since it depends on many factors, regression algorithms can aid in estimating this value.

**Ranking**. Some machine learning algorithms just rank objects by a number of features. Ranking is actively used to recommend movies in video streaming services or show the products that a customer might purchase with a high probability based on his or her previous search and purchase activities.



Figure 4.2 Dataset collection

**4.3.2 SPLITTING OF DATASET**

In machine learning ,any dataset is usually split into two : training data and test data. The output variable along with other variables are included in the training set . The model learns the data and tries to generate some pattern . The other part of the dataset serves as a test set to validate our model’s prediction. The scikit library has a function called train\_test\_split to divide our data . Test\_size is the parameter which gives us the percentage of data that should belong to the test set. Train\_size stores the remaining part as the training dataset , either of which should be specified. Random\_state acts as a random number generator . For our dataset , we split the training and testing set with 80 , 20 ratio the random state is passed as 0.

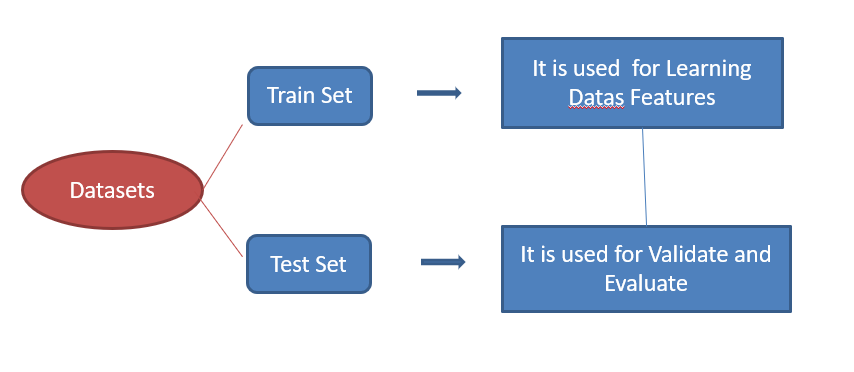


Figure 4.3 Splitting of Dataset

**4.3.3 DATASET PRE-PROCESSING MODULE**

Data pre-processing is a cleaning technique which is used to convert / transform the raw data into a clean and properly structured dataset suitable for further analysis. Data is usually collected and gathered from various sources , so it should be good enough and in some specific format before the model learns or gets trained with the data. This will help in achieving better and accurate results with valuable information. The basic steps in pre-processing involve filling up missing values and null values , getting rid of possible outliers and normalisation.

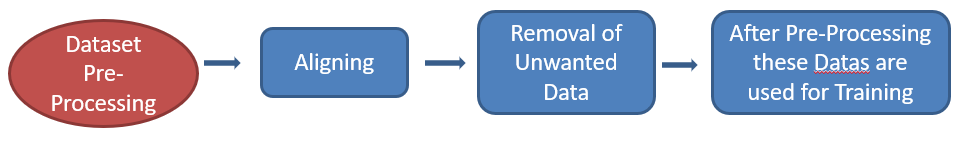


Figure 4.4 Dataset preprocessing

**4.3.4 TRAINING WITH ALGORITHM**

The Deep Neural Network is the regression machine learning algorithm to train the dataset consisting of abusive comments in social media. machine learning models require a lot of data in order for them to perform well. Usually, when training a machine learning model, one needs to collect a large, representative sample of data from a training set. Data from the training set can be as varied as a corpus of text, a collection of images, and data collected from individual users of a service. [Overfitting](https://en.wikipedia.org/wiki/Overfitting) is something to watch out for when training a machine learning model. Trained models derived from biased data can result in skewed or undesired predictions. [Algorithmic bias](https://en.wikipedia.org/wiki/Algorithmic_bias) is a potential result from data not fully prepared for training.

**CHAPTER 5**

**SOFTWARE DESCRIPTION**

**5.1 Jupyter notebook**

In this project the jupyter notebook is used as an IDE.

In this case, "notebook" or "notebook documents" denote documents that contain both code and rich text elements, such as figures, links, equations, ... Because of the mix of code and text elements, these documents are the ideal place to bring together an analysis description, and its results, as well as, they can be executed perform the data analysis in real time.

At some point, we all need to show our work. Most programming work is shared either as raw source code or as a co mpiled executable. The source code provides complete information, but in a way that’s more “tell” than “show.” The executable shows us what the software does, but even when shipped with the source code it can be difficult to grasp exactly how it works.

A notebook integrates code and its output into a single document that combines visualizations, narrative text, mathematical equations, and other rich media. In other words: it's a single document where you can run code, display the output, and also add explanations, formulas, charts, and make your work more transparent, understandable, repeatable, and shareable.

Using Notebooks is now a major part of the data science workflow at companies across the globe. If your goal is to work with data, using a Notebook will speed up your workflow and make it easier to communicate and share your results.

Imagine being able to view the code and execute it in the same UI, so that you could make changes to the code and view the results of those changes instantly, in real time? That’s just what [Jupyter Notebook](https://jupyter.org/) offers.

Jupyter Notebook was created to make it easier to show one’s programming work, and to let others join in. Jupyter Notebook allows you to combine code, comments, multimedia, and visualizations in an interactive document — called a notebook, naturally — that can be shared, re-used, and re-worked.

And because Jupyter Notebook runs via a web browser, the notebook itself could be hosted on your local machine or on a remote server

One major feature of the Jupyter notebook is the ability to display plots that are the output of running code cells. The IPython kernel is designed to work seamlessly with the matplotlib plotting library to provide this functionality. Specific plotting library integration is a feature of the kernel..

Each .**ipynb** file is one notebook, so each time you create a new notebook, a new  .**ipynb** file will be created.

Each **.ipynb** file is a text file that describes the contents of your notebook in a format called [JSON](https://en.wikipedia.org/wiki/JSON). Each cell and its contents, including image attachments that have been converted into strings of text, is listed therein along with some [metadata](https://ipython.org/ipython-doc/3/notebook/nbformat.html#metadata).

Jupyter Notebooks are a powerful way to write and iterate on your Python code for data analysis. Rather than writing and re-writing an entire program, you can write lines of code and run them one at a time. Then, if you need to make a change, you can go back and make your edit and rerun the program again, all in the same window.

Jupyter Notebook is built off of IPython*,* an interactive way of running Python code in the terminal using the REPL model (Read-Eval-Print-Loop). The IPython Kernel runs the computations and communicates with the Jupyter Notebook front-end interface. It also allows Jupyter Notebook to support multiple languages. Jupyter Notebooks extend IPython through additional features, like storing your code and output and allowing you to keep markdown notes.

Jupyter Notebook provides you with an easy-to-use, interactive data science environment across many programming languages that doesn't only work as an IDE, but also as a presentation or education tool.

**5.2** **Python**

Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is designed to be highly readable. It uses English keywords frequently where as other languages use punctuation, and it has fewer syntactical constructions than other languages.

Python is Interpreted − Python is processed at runtime by the interpreter. You do not need to compile your program before executing it. This is similar to PERL and PHP.

Python is Interactive − You can actually sit at a Python prompt and interact with the interpreter directly to write your programs.

Python is Object-Oriented − Python supports Object-Oriented style or technique of programming that encapsulates code within objects.

Python is a Beginner's Language − Python is a great language for the beginner-level programmers and supports the development of a wide range of applications from simple text processing to WWW browsers to games.

In this project python is used as programming language.

In technical terms, Python is an object-oriented, high-level programming language with integrated dynamic semantics primarily for web and app development. It is extremely attractive in the field of Rapid Application Development because it offers dynamic typing and dynamic binding options.

  Python is relatively simple, so it's easy to learn since it requires a unique syntax that focuses on readability. Developers can read and translate Python code much easier than other languages. In turn, this reduces the cost of program maintenance and development because it allows teams to work collaboratively without significant language and experience barriers.

  Additionally, Python supports the use of modules and packages, which means that programs can be designed in a modular style and code can be reused across a variety of projects. Once you've developed a module or package you need, it can be scaled for use in other projects, and it's easy to import or export these modules.

  One of the most promising benefits of Python is that both the standard library and the interpreter are available free of charge, in both binary and source form. There is no exclusivity either, as Python and all the necessary tools are available on all major platforms. Therefore, it is an enticing option for developers who don't want to worry about paying high development costs.

  If this description of Python over your head, don't worry. You'll understand it soon enough. What you need to take away from this section is that Python is a programming language used to develop software on the web and in app form, including mobile. It's relatively easy to learn, and the necessary tools are available to all free of charge.

**import pandas as pd**

import pandas as pd. Simply imports the library that current namespace, but rather than using the name pandas , it's instructed to use the name pd instead. This is just so you can dopd. whatever instead of having to type out pandas. whatever all the time if you just do importpandas.

**import numpy as np**

NumPy is an open-source numerical Python library. NumPy contains a multi-dimensional array and matrix data structures. It can be utilised to perform a number of mathematical operations on arrays such as trigonometric, statistical, and algebraic routines. NumPy **is** an extension of Numeric and Numarray.

**import Random**

import random imports the random module, which contains a variety of things to do with random number generation. Among these is the random() function, which generates random numbers between 0 and 1.

**import matplotlib.pyplot as plt**

Pyplot is a collection of functions in the popular visualization package Matplotlib. Its functions manipulate elements of a figure, such as creating a figure, creating a plotting area, plotting lines, adding plot labels, etc.

**import seaborn as sns**

Seaborn helps you explore and understand your data. Its plotting functions operate on dataframes and arrays containing whole datasets and internally perform the necessary semantic mapping and statistical aggregation to produce informative plots. Its dataset-oriented, declarative API lets you focus on what the different elements of your plots mean, rather than on the details of how to draw them.

**Sklearn**

Scikit-learn (Sklearn) is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistence interface in Python. It is a free software machine learning library for the Python programming language.

**sklearn.metrics**

Classification metrics. The sklearn. metrics module implements several loss, score, and utility functions to measure classification performance. Some metrics might require probability estimates of the positive class, confidence values, or binary decisions values.

**import roc\_curve**

ROC curves typically feature true positive rate on the Y axis, and false positive rate on the X axis. This means that the top left corner of the plot is the “ideal” point - a false positive rate of zero, and a true positive rate of one. This is not very realistic, but it does mean that a larger area under the curve (AUC) is usually better.

The steepness of ROC curves is also important, since it is ideal to maximize the true positive rate while minimizing the false positive rate.

ROC curves are typically used in binary classification to study the output of a classifier. In order to extend ROC curve and ROC area to multi-label classification, it is necessary to binarize the output. One ROC curve can be drawn per label, but one can also draw a ROC curve by considering each element of the label indicator matrix as a binary

**CHAPTER – 5**

**CONCLUSION AND FUTURE WORK**

**9.1 CONCLUSION**

In this project, we have successfully implemented a system of effective detection and identification of abusive comments over social media. The abusive comments dataset is trained using random forest classifier, a machine-learning algorithm. The present detection methods are inaccurate and inefficient. The system has provided an easy and efficient solution at very cost-effective approach.

Hence, The goal of our project is to filter tweets or comments using machine learning algorithms and to create an environment which is healthy for interaction and transfer of information between individuals.

We as the future generation are responsible for preventing the ongoing toxic environment in social media and create a healthy environment in social media.

**9.2 FUTURE WORK**

In the coming future, we review this application of abusive comment identification model to identification and detection of the abusive comment with more accuracy and efficiency. The application has good use in the social media. In the social media space, they have more chance to develop or convert this project in many ways. Thus, this project has an efficient scope in coming future where manual detection and prediction can be converted to computerized prediction in a cheap way.

**CHAPTER 10**

**REFERENCES**

1. Bounegru L, Gray J, Venturini T, Mauri M. A Field Guide to 'Fake News' and Other Information Disorders. A Field Guide to" Fake News" and Other Information Disorders: A Collection of Recipes for Those Who Love to Cook with Digital Methods, Public Data Lab, Amsterdam (2018). 2018.
2. Shrivastava, G., Kumar, P., Ojha, R. P., Srivastava, P. K., Mohan, S., & Srivastava, G. (2020). Defensive modeling of fake news through online social networks. IEEE Transactions on Computational Social Systems, 7(5), 1159-1167.
3. Kumar, A., Nayak, S., & Chandra, N.: Empirical Analysis of Supervised Machine Learning Techniques for Cyberbullying Detection. In: International Conference on Innovative Computing and Communications, pp. 223-230. Springer, Singapore (2019).
4. Mladenović, M., Ošmjanski, V., & Stanković, S. V. (2021). Cyber-aggression, cyberbullying, and cyber-grooming: a survey and research challenges. ACM Computing Surveys (CSUR), 54(1), 1-42.
5. Kumar, A., & Jaiswal, A. (2019). Swarm intelligence based optimal feature selection for enhanced predictive sentiment accuracy on twitter. Multimedia Tools and Applications, 78(20), 29529-29553.
6. Van Hee, C., Jacobs, G., Emmery, C., Desmet, B., Lefever, E., Verhoeven, B., ... & Hoste,

V. (2018). Automatic detection of cyberbullying in social media text. PloS one, 13(10), e0203794.

1. Hang, O. C., & Dahlan, H. M. (2019, December). Cyberbullying lexicon for social media. In 2019 6th International Conference on Research and Innovation in Information Systems (ICRIIS) (pp. 1-6). IEEE.
2. Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). Recent trends in deep learning based natural language processing. ieee Computational intelligenCe magazine, 13(3), 55-75.
3. Ballal N., Saritha S.K. (2020) A Study of Deep Learning in Text Analytics. In: Shukla R., Agrawal J., Sharma S., Chaudhari N., Shukla K. (eds) Social Networking and Computational Intelligence. Lecture Notes in Networks and Systems, vol 100. Springer, Singapore. https://doi.org/10.1007/978-981-15-2071-6\_16
4. Kumar, A., Srinivasan, K., Cheng, W. H., & Zomaya, A. Y. (2020). Hybrid context enriched deep learning model for fine-grained sentiment analysis in textual and visual semiotic modality social data. Information Processing & Management, 57(1), 102141.
5. Kumar, A., & Jaiswal, A. (2020). A Deep Swarm-Optimized Model for Leveraging Industrial Data Analytics in Cognitive Manufacturing. IEEE Transactions on Industrial Informatics, 17(4), 2938-2946.
6. Kumar, A., & Jaiswal, A. (2020). Systematic literature review of sentiment analysis on Twitter using soft computing techniques. Concurrency and Computation: Practice and Experience, 32(1), e5107.
7. Sangwan, S. R., & Bhatia, M. P. S. (2020). D-BullyRumbler: a safety rumble strip to resolve online denigration bullying using a hybrid filter-wrapper approach. Multimedia Systems, 1-17.
8. Brown, L. (2012). New Harvard Study Shows Why Social Media Is So Addictive for Many. [online] WTWH Marketing Lab. Available at: <http://marketing.wtwhmedia.com/new-harvard-study-shows-why-social-> media-is-so-addictive-for-many/ [Accessed 27 Jan. 2020].
9. https://coschedule.com/ Accessed on 14 July 2018.
10. Salawu S, He Y, Lumsden J (2017) Approaches to Automated Detection of Cyberbullying: A Survey. IEEE Transactions on Affective Computing (1):1-20.
11. Sarna, G., & Bhatia, M. P. S. (2020). Structure-Based Analysis of Different Categories of Cyberbullying in Dynamic Social Network. International Journal of Information Security and Privacy (IJISP), 14(3), 1-17.
12. Bounegru L, Gray J, Venturini T, Mauri M. A Field Guide to 'Fake News' and Other Information Disorders. A Field Guide to" Fake News" and Other Information Disorders: A Collection of Recipes for Those Who Love to Cook with Digital Methods, Public Data Lab, Amsterdam (2018). 2018.
13. Foody, M., Samara, M., & Carlbring, P.: A review of cyberbullying and suggestions for online psychological therapy. Internet Interventions 2(3), 235-242 (2015).
14. Campbell MA (2005) Cyber bullying: An old problem in a new guise?. Journal of Psychologists and Counsellors in Schools 15(1):68-76.
15. Barlett, C. P., Simmers, M. M., Roth, B., & Gentile, D. (2021). Comparing cyberbullying prevalence and process before and during the COVID-19 pandemic. The Journal of Social Psychology, 1-11.
16. Ybarra M (2010) Trends in technology-based sexual and non-sexual aggression over time and linkages to nontechnology aggression. National Summit on Interpersonal Violence and Abuse Across the Lifespan: Forging a Shared Agenda.
17. Cyberbullying, The National Crime Prevention. [http://www.ncpc.org/cyberbullying.](http://www.ncpc.org/cyberbullying) Accessed 10 March 2018.
18. Ngo, A. T., Tran, A. Q., Tran, B. X., Nguyen, L. H., Hoang, M. T., Nguyen, T. H. T., ... & Ho, C. S. (2021). Cyberbullying Among School Adolescents in an Urban Setting of a Developing Country: Experience, Coping Strategies, and Mediating Effects of Different Support on Psychological Well-Being. Frontiers in Psychology, 12, 930.
19. <http://www.ryanpatrickhalligan.org/>Accessed on 14 July 2018.
20. National Bullying Prevention Center. https://[www.pacer.org/bullying/.](http://www.pacer.org/bullying/) Accessed 26 July 2018.
21. Yuvaraj, N., Chang, V., Gobinathan, B., Pinagapani, A., Kannan, S., Dhiman, G., & Rajan, A. R. (2021). Automatic detection of cyberbullying using multi-feature based artificial intelligence with deep decision tree classification. Computers & Electrical Engineering, 92, 107186.
22. Farley, S., Coyne, I., & D’Cruz, P. (2021). Cyberbullying at work: Understanding the influence of technology. Concepts, Approaches and Methods, 233-263.
23. Chia, Z. L., Ptaszynski, M., Masui, F., Leliwa, G., & Wroczynski, M. (2021). Machine Learning and feature engineering-based study into sarcasm and irony classification with application to cyberbullying detection. Information Processing

& Management, 58(4), 102600.

1. Fang, Y., Yang, S., Zhao, B., & Huang, C. (2021). Cyberbullying detection in social networks using Bi-gru with self-attention mechanism. Information, 12(4), 171.
2. Kumar, R. (2021). Detection of Cyberbullying using Machine Learning. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(9), 656-661.
3. Bharti, S., Yadav, A. K., Kumar, M., & Yadav, D. (2021). Cyberbullying detection from tweets using deep learning. Kybernetes.
4. Mahat, M. (2021, March). Detecting Cyberbullying Across Multiple Social Media Platforms Using Deep Learning. In 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE) (pp. 299-301). IEEE.
5. Aggarwal, C. C. Neural Networks and Deep Learning: A Textbook. Springer 2018.
6. B. Kitchenham, S. Charters, (2007) Guidelines for performing Systematic Literature Reviews in Software Engineering, Tech. Rep. EBSE. 1 1–57.
7. Reynolds K, Kontostathis A, Edwards L (2011) Using machine learning to detect cyberbullying. InMachine learning and applications and workshops (ICMLA), 2011 10th International Conference, IEEE 2: 241-244.
8. Nahar V, Unankard S, Li X, Pang C (2012) Sentiment analysis for effective detection of cyber bullying. Asia-Pacific Web Conference, Springer, Berlin, Heidelberg: 767- 774.
9. Xu JM, Jun KS, Zhu X, Bellmore A (2012) Learning from bullying traces in social media. In Proceedings of the 2012 conference of the North American chapter of the association for computational linguistics: Human language technologies, Association for Computational Linguistics:656-666.
10. Kontostathis A, Reynolds K, Garron A, Edwards L (2013) Detecting cyberbullying: query terms and techniques. InProceedings of the 5th annual acm web science conference: 195-204.
11. Dadvar M, Trieschnigg D, Ordelman R, de Jong F (2013) Improving cyberbullying detection with user context. In European Conference on Information Retrieval, Springer, Berlin, Heidelberg: 693-696.
12. Sheeba JI, Vivekanandan K (2013) Low frequency keyword extraction with sentiment classification and cyberbully detection using fuzzy logic technique. In IEEE International Conference on Computational Intelligence and Computing Research (ICCIC): 1-5.
13. Nahar V, Al-Maskari S, Li X, Pang C (2014) Semi-supervised learning for cyberbullying detection in social networks. In Australasian Database Conference, Springer, Cham: 160-171.
14. Parime S, Suri V (2014) Cyberbullying detection and prevention: Data mining and psychological perspective. In Circuit, Power and Computing Technologies (ICCPCT), 2014 International Conference IEEE: 1541-1547.
15. Dadvar M, Trieschnigg D, de Jong F (2014) Experts and machines against bullies: A hybrid approach to detect cyberbullies. In Canadian Conference on Artificial Intelligence, Springer, Cham: 275-281.
16. Michalopoulos D, Mavridis I, Jankovic M (2014) GARS: Real-time system for identification, assessment and control of cyber grooming attacks. Computers & security 42:177-90.
17. Holt TJ, Turner MG, Exum ML (2014) The impact of self control and neighborhood disorder on bullying victimization. Journal of Criminal Justice 42(4):347-55.
18. Byrne S, Katz SJ, Lee T, Linz D, McIlrath M (2014) Peers, predators, and porn: Predicting parental underestimation of children's risky online experiences. Journal of Computer-Mediated Communication. 19(2):215-31.
19. Rafiq RI, Hosseinmardi H, Han R, Lv Q, Mishra S, Mattson SA (2015) Careful what you share in six seconds: Detecting cyberbullying instances in Vine. In Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ACM: 617-622.
20. Chavan VS, Shylaja SS (2015) Machine learning approach for detection of cyber- aggressive comments by peers on social media network. In Advances in computing, communications and informatics (ICACCI), 2015 International Conference on IEEE: 2354-2358.
21. Balci K, Salah AA (2015) Automatic analysis and identification of verbal aggression and abusive behaviors for online social games. Computers in Human Behavior 53:517-26.
22. Nandhini BS, Sheeba JI (2015) Online social network bullying detection using intelligence techniques. Procedia Computer Science 45:485-92.
23. Balakrishnan V (2015) Cyberbullying among young adults in Malaysia: The roles of gender, age and Internet frequency. Computers in Human Behavior 46:149-57.
24. Zhang X, Tong J, Vishwamitra N, Whittaker E, Mazer JP, Kowalski R, Hu H, Luo F, Macbeth J, Dillon E (2016) Cyberbullying Detection with a Pronunciation Based Convolutional Neural Network. In 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA): 740-745.
25. Zhao R, Zhou A, Mao K (2016) Automatic detection of cyberbullying on social networks based on bullying features. In Proceedings of the 17th international conference on distributed computing and networking: 43-48.
26. Hosseinmardi H, Rafiq RI, Han R, Lv Q, Mishra S (2016) Prediction of cyberbullying incidents in a media-based social network. In Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining: 186-192.
27. Gordeev D (2016) Detecting state of aggression in sentences using CNN. In International Conference on Speech and Computer, Springer, Cham: 240-245.
28. Hammer HL (2016) Automatic detection of hateful comments in online discussion. In International Conference on Industrial Networks and Intelligent Systems, Springer, Cham: 164-173.