International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET2310151

A Survey : Monitoring and Detecting Cyber Bullying Activities using Machine Learning Algorithms

Payal Budhe, Mrs. Dipalee Rane

Department of Computer Engineering D. Y. Patil College of Engg, Akurdi Pune, Maharashtra, India

ARTICLEINFO

Article History:

Accepted: 10 Feb 2023
Published: 23 Feb 2023

Publication Issue
Volume 10, Issue 1
January-February-2023

Page Number
374-383

ABSTRACT

The shrinking of the planet by technology is causing new age difficulties in youth culture. Technology surely has a lot of benefits, but it also has risks. It is where cyberbullying first started. Thus, there are many different types of cyberbullying. It might not necessarily involve pretending to be someone else or breaking into their online accounts. It also includes criticizing someone or spreading lies about them in an effort to cast doubt on them. Social media is widely used, making it incredibly easy for anyone to misuse this access. Cyberbullying is a serious issue today. It includes actions that harass, mislead, or defame someone. These violent behaviors are incredibly hazardous and can harm anyone quickly and severely. They appear on open discussion forums, social media sites, and other internet chat boards. A cyberbully is not always an anonymous person; they could be someone you know. The detection of online cyberbullying has grown in societal significance, research interest, and accessibility of open data. Even so, despite the continued rise in processing power and resource affordability, access limitations to high quality data constrain the use of cutting-edge methodologies. As a result, many recent studies use limited, heterogeneous datasets without fully assessing their usefulness. This study discusses effective techniques used to detect online abusive and bullying messages by merging natural language processing and machine learning algorithms with distinct features to analyze the accuracy levels of the algorithms.

Keywords: Cyber Bullying, Machine Learning Algorithm, Social Media.

I. INTRODUCTION

With more than four billion Internet users globally, the online world has tremendously influenced society and has become a necessary component of daily life. The world of today is totally dependent on technology, and young people are now leading modern lifestyle thanks to the internet. Cyberbullying is one of the main issues brought about by this rapid advancement in technology, which also has many disadvantages. The internet has developed as a multiple tool that has significantly facilitated our daily lives. However, the internet has also provided a platform for a variety of undesirable activities, such as cyberbullying.

A. CYBERBULLYING

Cyberbullying, also known as cyber harassment, is the use of certain internet gadgets to threaten, bully, harass, or intimidate someone. This also goes by the name of online bullying. Bullying via a digital platform, media, or gadget is cyberbullying. Cyberbullying doesn't always involve posing someone else or hacking into someone's account or profile. Yet there are a lot of different ways that cyberbullying can happen. Cyberbullying is the act of distributing false information about another person online, including through text messages sent by SMS, online chat rooms, game forums, social networking sites, and online chat. It can be viewed on a variety of digital devices, including tablets, smartphones, and laptops. When offensive, harmful, or inappropriate content is sent, uploaded, or shared using various digital tools, it is referred to as cyberbullying. Cyberbullying has become a widespread issue because everyone uses social networking sites today, and it's easy to abuse this access. Embarrassing, blackmailing, disparaging, manipulating, or harassing activities are included in this. Such acts of aggression simply and unfavourably cause a person to suffer severe harm.

B. CYBERBULLYING TYPES ACCORDING TO THE LITERATURE, THERE ARE 12 TYPES OF CYBERBULLYING [24]:

1) Flaming: Starting a fight online.
2) Harassment: Sending insulting messages frequently.
3) Cyberstalking: Sending intimidating messages to the victim, which causes fear.
4) Masquerade: The bully pretends to be someone else.
5) Trolling: Posting controversial comments to upset other members on the online platform.
6) Denigration: Negative gossip about another person.
7) Outing: Posting personal information about someone in public forums.
8) Exclusion: When a social group deliberately excludes someone.
9) Catfishing: Creating a fake profile using someone else's information.
10) Dissing: Posting information about someone to hurt them or defame them.
11) Trickery: Tricking someone to share their secrets or personal information.
12) Fraping: Using someone else's online account to post inappropriate content and tricking others into believing that the account owner posted them.

C. COUNTERMEASURES BY SOCIAL MEDIA

Users can report bullying on social networking sites like Facebook and Twitter, which promote a safe environment online. These include specifying the intended audience, blocking specific users, and recognizing and banning people who behave badly. Despite the fact that they are incredibly important, these techniques are reactive in nature and only apply after the victim has already been harmed. By the time someone reports the offensive post and the required action is taken by the authority, many users may have already read it, having already experienced the previously mentioned harmful effects. We therefore need an automated system that can quickly and accurately identify cyberbullying behaviour.

D. FEATURE TYPES USED IN CYBERBULLYING PREDICTION

TABLE 1.
SUMMARY OF CONTENT BASED FEATURE TYPES USED IN CYBERBULLYING

	Content Based Features					
Paper	**BoW**	**SG**	**PF**	**CB**	**SF**	**PR**
1	√	√	×	×	×	√
2	√	×	√	√	×	√
3	×	×	√	×	×	√

4	√	×	√	×	×	×
5	√	×	×	×	√	×
6	√	×	√	×	×	×
7	√	×	√	×	√	×
8	√	×	×	×	×	×
9	√	×	√	×	×	×
10	√	×	√	√	√	√
11	√	×	√	√	×	×
12	√	×	×	×	√	×
13	√	×	×	√	×	×
14	√	×	√	×	√	×

BoW - bag of words, SG - skip gram, PF - profanity features, SF - sentiment features, PR – pronouns

TABLE 2.

SUMMARY OF PROFILE BASED FEATURE TYPES USED IN CYBERBULLYING

	Profile Based Features			
Paper	**DF**	**FCF**	**TSF**	**LOCF**
1	×	×	×	×
2	×	×	×	×
3	√	×	×	×
4	×	×	×	×
5	×	×	×	×
6	×	√	×	×
7	×	×	×	×
8	×	×	×	×
9	×	×	×	×
10	√	√	×	×
11	×	×	×	×
12	×	×	×	×
13	×	×	√	√
14	×	×	×	×

DF - demographic features , FCF - friends or follower count features, TSF - timestamp features, LOCF - location of post feature

1. Bag of Words : It is a simplified representation used in information retrieval and natural language processing. 2. Skip Gram : An unsupervised learning technique is used to identify the terms that are most associated with a particular word.

3. Profanity Features: It should always be used even if only to capture and omit the mist offensive word.

4. Sentiment Features: It is the combination an action of belief and emotions that explain for example positive, negative, happy, sad etc.

5. Pronouns: A pronoun is defined as a word or phrase that is used as a substitution for a noun.

6. Demographic Features: The term "demographic features" refers to differences in a society based on factors such as gender, age, occupation, education level, religion, ethnicity, income, marital status, and a variety of other aspects of the population.

7. Friends and Followers count Feature: The difference between friends and followers is how much access people have to your profile and content. Social media friend is a two-way relationship. When you accept to be someone's friend, you see each other's posts. However, following is a one-way relationship. You see content from the person you follow, but they don't see yours.

8. Timestamp Features : When data is added, deleted, modified, or sent, a timestamp is a time that is associated with the file, log, or notice.

9. Location Post Features: A location is the place where a particular point or object exists.

II. RELATED WORK

Several studies have been conducted to identify cyberbullying.

In [23], This article introduces a brand-new Bully Net architecture for locating bullies on the Twitter social network. In order to create an SN based on bullying tendencies, researchers conducted in-depth research on mining SNs for a better understanding of the interactions between users in social media. They found that by creating conversations focused on environment as well as content, they could successfully pinpoint the feelings and actions that cause bullying. During the experimental investigation, the examination of their suggested centrality metrics to recognize bullies from SN, they were able to identify bullies for a variety of scenarios with about 80% accuracy and 81% precision. In [18] this research, researchers suggested a detection architecture for cyberbullying to address the issue. They talked about the data architecture for hate speech on Twitter and personal attacks on Wikipedia. Given that tweets containing hate speech typically contained cursing, which made it simple to identify, natural language processing techniques for this type of speech were successful with accuracy rates of over 90% utilizing fundamental machine learning algorithms. Because of this, using BoW and Tf-Idf models rather than Word embeddings models produces better results. Although the three feature selection approaches worked similarly, it was challenging to identify personal assaults using the same model because the comments lacked a lot of learnable sentiment.

In [22], Haider et al. discuss a study on the identification of multilingual cyberbullying. They discovered that the majority of work in this field is done in English, thus they tried to identify cyberbullying in Arabic. They employed ML learning techniques to identify cyberbullying in their work. 32K tweets made up their dataset, and 1800 of those were bullying-related. To identify cyberbullying, they utilized the Support Vector Machine (SVM) and Naïve Bayes methods, and they received F1 scores of 92% and 90%, respectively.

In [20] this study, researchers developed two ensemble-based voting algorithms to identify sentences that are offensive or not. Every ML algorithm and ensemble technique that was used independently has been outperformed by our suggested model. For the twitter extracted dataset, they had the greatest 96% accuracy. The performance of their model will be evaluated in the future using a variety of diverse datasets, as well as some private datasets. Finally, there are many other types of cyberbullying, including harassment, flame, denigration, impersonation, racism, sexism, etc.

In [16] this paper, the issue of detecting cyberbullying was addressed by the sequential hypothesis testing methodology. More specifically, the objective is to choose when to stop extracting and evaluating features from the message and make a decision. Each communication can be classified into one of two classes (i.e., cyberbullying or normal). In order to achieve this, an optimization function was created in terms of the average cost of the classification technique and the cost of features, and the best possible outcome was found.

III. METHODOLOGY

These actions are required to identify cyberbullying. [26]:

- Open the Kaggle repository and load the dataset.

-Pre-process the dataset by cleaning the text, tokenizing, stemming, lemmatizing, and removing stop words. After text cleaning, linguistic techniques were utilized to examine the pattern of offensive comments.

-The dataset was then divided into training and test sets. Train different algorithms on the dataset. Utilize the testing dataset and a variety of metrics to evaluate the effectiveness of the algorithms.

I. Dataset: The first step in identifying cyberbullying is gathering data sets from various online networks. On social networking and media websites, user comments, posts, images, and videos frequently create data sets for cyberbullying. Accessing tweets on Twitter is made simple by using the Twitter API. In addition to pre-made datasets from websites like kaggle.com, data from websites like YouTube, Facebook, Myspace, Instagram, and others is also used for the identification of cyberbullying.

II. Pre-processing: The next step is data pre-processing, which is used to change the data set so that it only contains relevant data. Prior to tokenization and lemmatization, special characters, stop words, and white spaces are removed from the data as part of data pre-processing. We may now structure a data collection using a variety of additional techniques as well.

a. Tokenization: Tokenization is the process of dividing a text document into tokens, which are very small pieces of text. Words, word fragments, or simple characters like punctuation can all be considered tokens. Word, character, and sub word (n-gram characters) tokenization are the three main types of tokenization. Word The most used tokenization algorithm is tokenization.

b. Stemming: Humans desire to reduce words to their root or basic form after tokenizing, or breaking sentences into words. In effect, stemming is exactly what is meant by this. Stemming is the practice of combining words with similar meanings into their "stem" or "root" versions.

c. Lemmatization: The method of lemmatization involves evaluating a word's several inflected forms as a single entity. Lemmatization gives the words additional context, just like stemming does. It thus connects words with similar meanings.

d. Stopword removal: The most frequent words in any language that have no meaning are called stop words, and natural language processing typically ignores them. Stop words in English include "a," "and," "the," and "of." Stop words are frequently eliminated from texts in natural language processing before they are processed for analysis. This is done to simplify the content and exclude unnecessary information. Data classification is the third and last step in the detection of cyberbullying. The information is divided into instances of positive or negative cyberbullying, i.e., information that most definitely contains information about cyberbullying against information that doesn't significantly includes information about cyberbullying. Before identifying input data, classification algorithms must first predict the label of an input using a training set of labelled samples. For data classification, a variety of algorithms and techniques can be utilized.

This module applies a variety of techniques, including Binary Classification, Fine-Grained Classification, Decision Tree, Random Forest, Support Vector Machine, and Naïve Bayes etc. to identify the bullying-related text and message. To identify the most accurate classifier for a certain public cyberbullying dataset.

Algorithms:

1) Binary Classification:

If the classification problem has only two possible outcomes, then it is called as Binary Classifier. Binary classification is a supervised learning algorithm. Concentrate your study on binary categorization, where the outcome mostly depends on whether a message contains bullying or not. Binary Classification primarily focuses on text-based feature. The features of an article or report that are not part of the primary text are referred to as "text-based features" [table of contents, glossary, index, headers, bold text, sidebars, and pictures]. The implementation of binary classification to distinguish bullying and non-bullying traces[15].

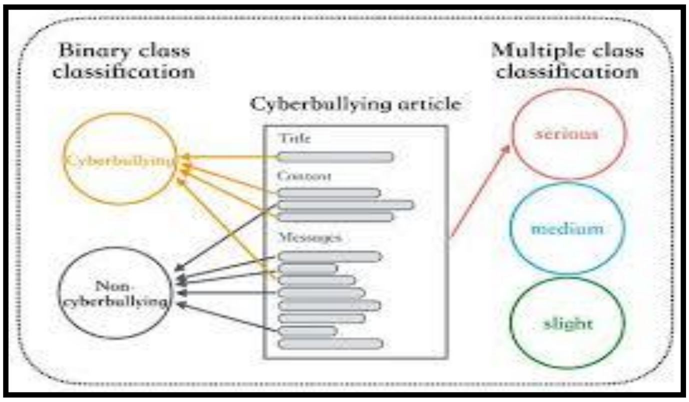

Fig 1. Binary Classification and Multi Class Classification[27]

2) Fine-Grained Classification:

By examining bullying traces or the responses to a bullying incident, we can extend this binary technique with more precise categorization tasks. Based on (bully) Twitter data that was keyword-retrieved, they separated two tasks: (1) a role-labelling task in which person-mention roles were distinguished using semantic role labelling; (2) the addition of sentiment to detect teasing. The difficulty of fine-grained classification of bullying roles (harasser, bystander assistant, bystander defender, victim) and types (curse, defamation, defence, encouragement, insult, sexual, threat) with straightforward Bag of Words and sentiment traits. The text-based feature is another focus of the fine-grained approach. When compared to a baseline of curse words and word n-grams, a multifeatured model had the lowest error rates for all labels for both bullying type and role classification [15]. Their sociolinguistic model for bullying classification also includes fine-grained information about bullying incidents. They partially rely on Twitter's social interactions to uncover latent text types and roles. As a result, it relates to the subcategory of work that uses the network where the data is collected meta-data. They can reach a situation where the dangers are detected by employing this model.

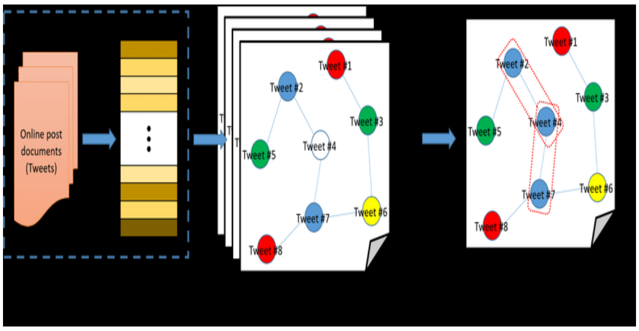

Fig 2. Fine Grained Classification[27]

3) Metadata Feature:

Metadata is simply data about data. It means it is a description and context of the data. It helps to organize, find and understand data. Using this data together with a classifier trained on LDA (Linear Discriminant Analysis) and weighted Tf-Idf features, Nahar et al. (Author) were the first to apply this approach to identify bullies and victims in the datasets [15]. This task's use of a graph to visualize prospective bullies and their connections is notable but less well liked. BoW characteristics are the most effective in Metadata Feature. By using this model, they can achieve a state where the information of cyberbully commented people, identified the rumors tweet, cyberbully tweeted people in formation .

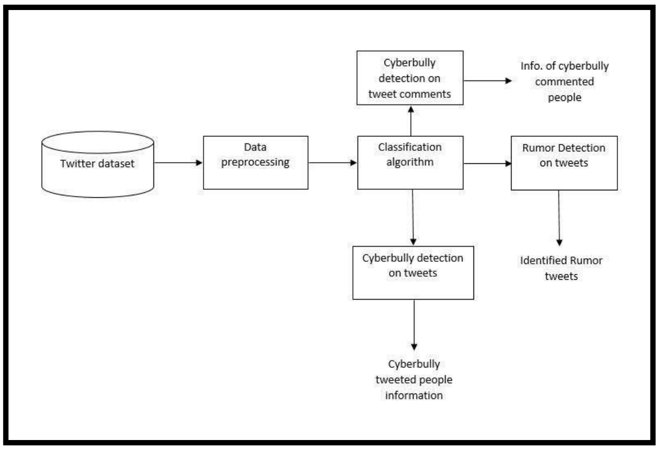

Fig 3. Metadata feature representation[27]

4) SVM (Support Vector Machine):

One of the most effective and versatile classification algorithms is the support vector machine. Finding the optimal separating hyperplane that maximizes the training data margin is its goal. The classifier is initially trained using labelled data before being used to categorize the data to test accuracy. To use the data to train the classifier, the data must be processed. Instead of giving the feature vector, similarity between data points can be shown using kernel functions. SVM can be used with numerous kernels, including:

- RBF kernel (Radial basis function)
- Linear kernel
- Gaussian kernel

The linear kernel, a subset of the RBF kernel, functions most effectively when there are a lot of features. The Myspace, Kongregate, and Slashdot data sets were used with the linear kernel. The content analysts have access to the datasets. The datasets include manually labelled data from, which serves as a reference dataset. The dataset includes three different social networking sites i.e Slashdot, Kongregate and MySpace. User messages from chat logs are provided by Kongregate, an online gaming portal. Due to the inherent frustration of playing online games and the use of language to communicate with opponents, aggression is frequently seen in the posts. Slashdot is a social networking platform where users can broadcast comments to other users during discussions. MySpace is a well-known social networking platform. Datasets are XML files that include and describe a conversation thread with several posts. Every post was isolated as a separate data piece. Each data element is treated as a separate document and indexed using the inverted file index, with a weight assigned to each unique phrase . A linear kernel, tenfold cross-validation, and LibSVM are used, and the results show a false positive rate of 28 in 294 instances and a false negative rate of 12 in 10184 instances. In this case, the weighted TF-IDF model was used. By oversampling the training cases, the training was enhanced. Bigrams should be used with the SVM linear kernel since bigrams produce an accuracy of 83.3%, while using unigrams only produces an accuracy of 79.6%. This conclusion was reached when the aforementioned hypothesis was tested using data from the Twitter corpus (1762 tweets, 39% of which had bullying traces). Considering all of this data, researchers discover that the linear SVM in conjunction with bigrams offers the highest level of accuracy[26].

5) J 48

A single-variable decision tree algorithm is used. The objective of this method is to categories data by processing it through a decision tree. On the basis of previously categorized dummy data sets, the decision is made as to which class the provided data belongs to. The decision tree's internal nodes describe multiple qualities, and the branches show us every possible value for each attribute. The final data classification is provided by the leaf nodes. It is believed that the feature with the maximum information gain can provide us with the most data for categorization. These are the features that will select for classification in order to be easier and more efficient.

The data is assigned to the relevant classifications after being divided into occurrences of bullying and non bullying. As a result, the classifier can successfully separate the tweets that have as input[26].

6) Naive Bayes:

The Naïve Bayes family of classifiers performs simple conditional probabilistic classification by applying the Bayes theorem with naive independence assumptions among the various features. Here, a very basic document representation typically a bag of words is used. Words that are crucial to the text's classification and hence essential to understanding its meaning are taken into account and given weight in keeping with their significance, or severity, in this situation. For example, "faggof" would be given a higher weight than "bitch" because the former is harsh and discriminatory toward women. The most likely class, or maximum posterior class, in this instance is whether bullying is occurring or not. The data corpus obtained for testing purposes is the same as the one used for J48. In this instance, accounting for both textual and social aspects, a true positive rate of 0.723 was achieved. The rate was 0.584 when social factors were ignored[26].

7) Random Forest:

Multiple "trees" are used in the categorization process by the Random Forest Algorithm. Every tree generates a classifier, and these classifiers cast votes to determine which algorithm receives the most votes. The immense amount of datasets are then characterized using this classification calculation. In the presented model, they can see that the natural language tool kit Library helps in our ability to interpret the meaning of the given sentences. The system is informed about the significance of the terms used in the dataset by the sub library, such as Wordnet. Numerous variables, including synonyms, age, location, gender, hashtags, and sarcasm, are taken into account during implementation. The connections between language-based keywords like hashtags, keywords, and synonyms are independent from the connections between a user's personal information, such as age, gender, and location. A search between these sets of information is possible. They discover that because random forest uses many trees to draw analytical conclusions, the True Positive rate, which reflects the instances of accurate predictions, and the Recall value, which is the probability of correct predictions, are both extremely high. They can also see that the Precision (Positive Predictive Value) is close to 0.800, indicating that the model is more accurate at predicting security irregularities and privacy threats when many factors are taken into account.

8) Social Signed Networks :

Examine the problem of cyberbullying on social media in an effort to respond to the following research question. Can the context of tweets (conversations) help Twitter better identify cyberbullying? The instinct is that each tweet should be evaluated based on both its content and the context in which it is found. A series of tweets between two or more people exchanging information about a certain topic is what experts refer to as a conversation in this context. Thus, there are three components to the solution. A discussion graph is first created for each interaction based on the attitude and hateful language used in the tweets. Second, by combining all discussion graphs and determining the bullying score for each user pair, researchers create an SSN they call the bullying SN. When only using positive connections, negative connections can reveal information that would normally go unnoticed. Finally, they propose an attitude and merit (A&M) centrality measure [23] to detect users who bully others on the SN.

Performance Metrics :

A. Accuracy : The ratio of the number of bully users to the overall number of bullies is the accuracy metric. It Doesn't perform well with imbalanced dataset[23].

Accuracy = # of detected bullies/total number of bullies.

B. Precision and Recall :

In binary classification tasks, evaluation metrics such as recall and precision are used. Recall is a measure of completeness, while precision is a measure of exactness. They are defined as follows[23] :

Precision = # of true bullies detected/ total number of detected users.

Recall = # of true bullies detected/total number of true bullies.

C. F1 Measure :

The harmonic mean of recall and precision is known as the F1 measure. F1 has a range of [0, 1]. It measures how reliable it is and how many bullies are accurately identified. Mathematically, it can be expressed as[23] . F1=2*Precision*Recall/Precision + Recall.

TABLE 3.
PERFORMANCE EVALUATION METRICS

Classifiers	**Accuracy**	**Precision**	**Recall**	**F1**
SVM	89.75%	0.885	0.896	0.886
J 48	89.71%	0.890	0.901	0.886
Naïve Bayes'	75.52%	0.858	0.802	0.791
Random Forest	86.57%	0.898	0.907	0.864
Signed Networks	73.60%	0.813	0.776	0.794

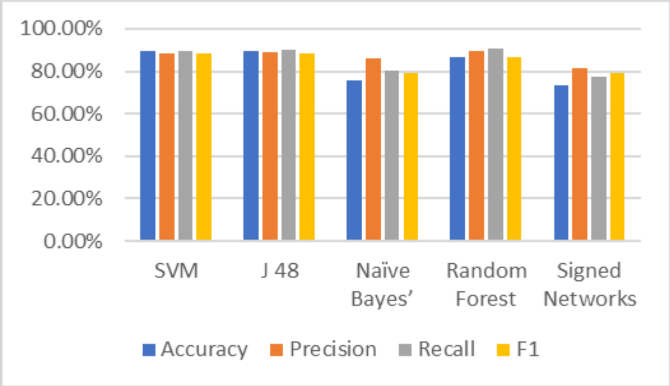

Fig 4. Classification Report

IV. CONCLUSION

This study analyzed a collection of studies on the topic of utilizing different algorithms to identify aggressive conduct on social networking websites. Additionally, a summary of different categories of discriminative traits that were utilized to identify cyberbullying on online social networking sites was provided. Support Vector Machine was proved to be the best classifier with Accuracy 89.75% and F-measure 0.886. The development of information and networking technologies has resulted in responses to online communication that are both positive and terrible, as well as ugly. These reactions are frequently misused and have left innocent people with lifelong emotional trauma, which frequently results in depression and suicide. They were unable to speak up to ask for assistance from various organizations or family members.

V. REFERENCES

[1]. Chavan, V.S. and S. Shylaja. 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. 2015. IEEE.

[2]. Chen, Y., et al. Detecting Offensive Language in social media to Protect Adolescent Online Safety. in Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Conference on Social Computing (SocialCom). 2012. IEEE

[3]. Dadvar, M., et al., Improved cyberbullying detection using gender information. 2012.

[4]. Dinakar, K., R. Reichart, and H. Lieberman, Modeling the detection of Textual Cyberbullying. 2011.

[5]. Van Hee, C., et al. Detection and fine-grained classification of cyberbullying events. in International Conference Recent Advances in Natural Language Processing (RANLP). 2015.

[6]. Hosseinmardi, H., et al., Detection of cyberbullying incidents on the Instagram social network. arXiv preprint arXiv:1503.03909, 2015

[7]. Kontostathis, A., et al. Detecting cyberbullying: query terms and techniques. in Proceedings of the 5th annual acm web science conference. 2013. ACM.

[8]. Sanchez, H. and S. Kumar, Twitter bullying detection. UCSC ISM245 Data Mining course report, 2011.

[9]. Zhao, R., A. Zhou, and K. Mao. Automatic detection of cyberbullying on social networks based on bullying features. in Proceedings of the 17th International Conference on Distributed Computing and Networking. 2016. ACM.

[10]. Squicciarini, A., et al. Identification and characterization of cyberbullying dynamics in an online social network. in Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015. 2015. ACM.

[11]. Reynolds, K., A. Kontostathis, and L. Edwards. Using machine learning to detect cyberbullying. in Machine Learning and Applications and Workshops (ICMLA), 2011 10th International Conference on. 2011. IEEE.

[12]. Yin, D., et al., Detection of harassment on web 2.0. Proceedings of the Content Analysis in the WEB, 2009.

[13]. Xu, J.-M., et al. 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. 2012. Association for Computational Linguistics.

[14]. Galán-García, P., et al. Supervised Machine Learning for the Detection of Troll Profiles in Twitter Social Network: Application to a Real Case of Cyberbullying. in International Joint Conference SOCO'13-CISIS'13-ICEUTE'13. 2014. Springer.

[15]. Chris Emmery,Ben Verhoeven, Guy De Pauw, Gilles Jacobs, Cynthia Van Hee, Els Lefever, Bart Desmet, Ve ́ronique Hoste, Walter Daelemans, Current limitations in cyberbullying detection: On evaluation criteria, reproducibility, and data scarcity,Springer2020.

[16]. Saloni Mahesh Kargutkar, Prof. Vidya Chitre , A Study of Cyberbullying Detection Using Machine Learning Techniques , IEEE Xplore 2020.

[17]. Mohammed Ali Al-garadi1,Mohammad Rashid Hussain, Nawsher Khan, Ghulam, Murtaza, Henry Friday Nweke1, Ihsan Ali, Ghulam Mujtaba,Haruna Chiroma, Hasan Ali Khattak and Abdullah Gani," Predicting Cyberbullying on Social Media in the Big Data Era Using Machine Learning Algorithms: Review of Literature and Open Challenges ", IEEE 2019.

[18]. Monirah A., Al-Ajlan, Mourad Ykhlef, Optimized Twitter Cyberbullying Detection based on Deep Learning, 978-1-5386-4110-1, IEEE-2018.

[19]. N. M. Zainudin, K. H. Zainal, N. A. Hasbullah, N. A. Wahab, and S. Ramli, "A review on cyberbullying in Malaysia from digital forensic Vandana Nanda Kumar, Binsu C, Kovoor, Sreeja M.U., Cyber - Bullying Revelation in Twitter Data using Naïve-Bayes Classifier Algorithm International Journal of Advanced Research in Computer Science. Volume 9, No. Jan-Feb 2018.

[20]. Semiu Salawu, Yulan He, and Joanna Lumsden, Approaches to Automated Detection of Cyberbullying: A Survey , IEEE Transaction 2017.

[21]. Just in W. P at chin, Summary of Our Cyber bullying Research (2007 -2019), Cyberbullying Research Centre, July 10, 2019.

[22]. Rohit Pawar, Rajeev R. Raje, Multilingual Cyber bullying Detection System.,IEEE 2019.

[23]. Aparna Sankaran Srinath , Hannah Johnson, Gaby G. Dagher , and Min Long, BullyNet: Unmasking Cyberbullies on Social Networks,IEEE2021.

[24]. Farhan Bashir Shaikh, Mobashar Rehman, and Aamir aamin, Cyberbullying: A Systematic Literature Review to Identify the Factors Impelling University Students Towards Cyberbullying., IEEE 2020.

[25]. Bandeh Ali Talpur, Declan O'Sullivan, Cyberbullying severity detection: A machine learning approach ,Plos one 2020.

[26]. Rekha Sugandhi, Anurag Pande, Siddhant Chawla, Abhishek Agrawal, Husen Bhagat, Methods for Detection of Cyberbullying: A Survey, 2015 15th International Conference on ISDA

[27]. https://www.bing.com/image/search?q=cyberbullying+detection+diagramform=HRDSC2&first=1&tsc=ImageHoverTitle

Cite this article as :

Payal Budhe, Mrs. Dipalee Rane , "A Survey on Monitoring and Detecting Cyber Bullying Activities using Machine Learning Algorithms", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 10 Issue 1, pp. 374-383, January-February 2023. Available at doi : https://doi.org/10.32628/IJSRSET2310151

Journal URL : https://ijsrset.com/IJSRSET2310151