ISSN 2320-9984

Volume 10 Issue 2 | June 2022 ISSN: 2320-9984 (Online)

International Journal of Modern Engineering & Management Research

Website: www.ijmemr.org

Cyberbullying Detection in Social Media Using Supervised ML and NLP Techniques

Indu Sai Raman
M.Tech. Research Scholar
Computer Science and Engineering
Takshshila Institute of Engineering and Technology
Jabalpur, (M.P.) India
Email: indusairaman26@gmail.com

Akshat Khaskalam
Assistant Professor
Department of Mechanical Engineering
Takshshila Institute of Engineering and Technology
Jabalpur, (M.P.) India
Email: akshatkhaskalam@takshshila.org

Abstract— *From the era internet originated into survival, the social networking epoch has been sprouting. In the opening phase, no one believed that the internet will be desired as a cloud of great services such as social-networking. Now, can say that social networking and virtual application are non-separable and obligate to a portion of humans' life. Numerous people from various age clusters spend most of their time on websites. Even though the statistics that people are passionately associated through social media, these amenities convey along enormous dangers with them, such as cyber attacks that comprises cyberbullying. Along with the increase in social networks, cyberbullying also increases day-by-day. To recognize word resemblances in the twitters made by bullies and with the usage of machine learning (ML)a cyberbullying detection system is developed. However, various social media bullying detection methodology has been applied, but several of them were word-based. Under this motivation and background, the proposed system developed a cyberbullying detection in social media using a supervised ML and natural-language pre-processing (NLP) techniques. Though, the system is designed to discover cyberbullying in social-media, the system used twitter text and comments has input data, the preprocessing is performed with NLP techniques to enhance the detection result. The system efficiently performed feature extraction with count vectorization and the system used Naïve Bayes and Decision tree for the classification. Therefore, efficiency and effectiveness of the planned method is estimated in terms of performance metrics such as F1-score, precision, recall and accuracy.*

Keywords:— *Cyberbullying, NLP, Machine learning, Naïve Bayes, Decision tree.*

1. INTRODUCTION

The social-networking spots are spreading extensively today for numerous activities such as networking, entertainment. Today, social-networking sites were a stop for various reasons to billion peoples. The social media network required the consent of all participating people. Interactive with people is no exclusion, as expertise has been developed, the interaction among people with a broader method and new measurement to communicate was given. Though, the system provides innovate idea in communication, many of them were illegally using these populations. Due to this many teenagers are in receipt of bullied. Bullies use various services such as Face-book-Mail, Twitter[1].

One of the most frequently happening internet abuse is cyberbullying and also it was also a serious social issues particularly for teenagers. Due to this, more researchers were devoting on the prevention and detection of cyberbullying specifically in social media. Cyberbullying was not just constrained to create a fake identity and posting (or) publishing some embarrassing videos or photo, unpleasant rumors about a person may create threats to them. Cyberbullying impact on social media were horrifying, sometimes it leads to death of unfortunate victims.

Thus, a comprehensive solution is needed for this challenges. The cyberbullying needs to stop. With the support of machine learning, the cyberbullying was detected and prevented. The implementation should be done with different perspective. The significant growth of internet equipment, social-media platforms like Facebook and twitter has developed popular and played vital role in altering human life. In specific, social media has been incorporated with human's daily activities such as entertainment, business, education, e-government and business. Accordingly, social networking effects were proposed to surpass 3.02 billion dynamic social media handlers every month. This ratio will justification around one third of earth's populace. However, between various prevailing social networks, Twitter was a crucial platform and has a significant data source for researchers. Twitter was a widespread public micro blogging structure functioning in present. Characterized with its short message limit and un- filtered feed, the use of twitter has been increased rapidly, on an average of 500

2. LITERATURE REVIEW

The various machine learning has been implemented in the detection of cyber bullying. Cyber-bullying mainly happens in the social media platforms. Therefore; the proposed system used machine learning approach for the detection of cyber-bullying. In order to implement the system successfully in the prevention and detection of cyber bullying the existing system implemented in the cyber bullying were reviewed to find the strength and weakness for implementing the proposed system.

Cyber Bullying Detection Using Machine Learning:

In recent times, the usage of internet and electronic devices has been increased, especially with teenager. But this growth brings a great effect to human life, some people use this for malicious purpose, one of those activities is the cyberbullying. It was a form of bullying done over electronic means in order to harm and insult others. Though, many researchers have proposed strategies and solution for detecting the cyber bullying, but the sarcasm has been still happening. Therefore, the study[1] the proposed and previous work in the detection along with the elements included in sarcasm. The system SVM classifier for the detection.

Cyberbullying is the malicious activities which troubles others. It mostly appears in the social media platform. It happens in the text format. Most of the existing system has approached this issue with conventional ML techniques and majority of the implemented system in these studies were adaptable to a single social network at a time. Therefore, the proposed system[2] aimed to examine the developed model performance with a new dataset along with this the investigation was performed in platform. These systems showed better results in the prediction and detection of the cyberbullying.

3. PROBLEM STATEMENT

The cyber bullying was increasing rapidly, thoug`h various systems has been implemented in the detection of cyber bullying still the lacks in some points. This section clearly states the weakness of the existing system in cyberbullying detection. This leads to implement the proposed in an effective manner for detecting the cyber bullying in social media. The following section states the problems in existing system.

Research Gap:

The growth of social-media, specifically twitter, raised numerous problems due to the misinterpretation concerning freedom of communication. This is one of the major issued that affects both societies and distinct victims. Numerous efforts have been proposed in the works to preclude or mitigate cyber bullying; moreover, these efforts rely on the connections of victims. Therefore, cyberbullying discovery without the involvement of victims are necessary. Based on this, various system have been implemented, though the system showed its efficiency in the detection still it fails in some points such as, The system lacks in handling large volume of data[9], also shows theoretical limits, the classification results are incorrect, accuracy range in prediction was very low. In this, LSF (lexical syntactic features) cannot handle huge dataset in the prediction[10]. Automated sentiment analysis tools performs great function in the text analyzing for attitude and opining, but system failed to achieve this advantage[11]. In this system, crowd flower workforce shows some limitations and were difficult to manage[12]. Similarly, machine learning requires huge data set for training, it should be unbiased/ inclusive and of decent quality[13]. Large time taken for the generation of new dataset. The timely detection is main, but systems fail to detect the cyber-bullying in a correct time.

4. EXISTING SYSTEM AND PROPOSED SYSTEM

This section deliberates the proposed work of the system. The system used Navies Bayes for comparison and the proposed system was performed with Decision tree for cyber bulling detection. The system architecture was explained in the following section.

4.1 Proposed System Framework:

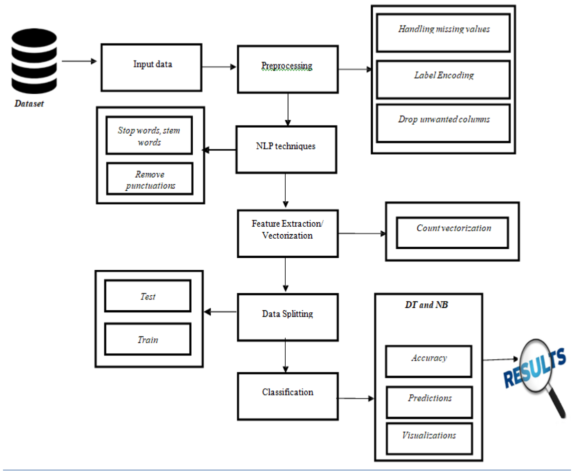
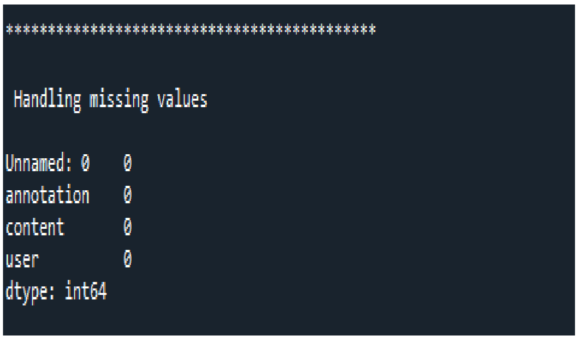


Figure 1: System Architecture

The system introduced cyberbullying detection in social media using supervised machine learning and natural language pre-processing techniques. Though the system was proposed to detect the cyber- bulling in social media, the dataset was selected from twitter and the pre-processing was performed by labeling encoding, handling missing values and dropping unwanted columns and the NLP techniques were implemented for pre-processing by removing punctuations, stop words and stem words. The feature extraction was performed to extract the relevant features, the count vectorization was utilized for the feature extraction. After, the feature selection process, the data splitting was performed, the system randomly splits the data as testing and training. Then, the classification was performed with decision tree (DT) and Naïve Bayes(NB). The system proposed decision tree technique and the Naïve Bayes was the existing methods utilized for comparison analysis.



```
********************************************


 Handling missing values

Unnamed: 0    0
annotation    0
content       0
user          0
dtype: int64
```

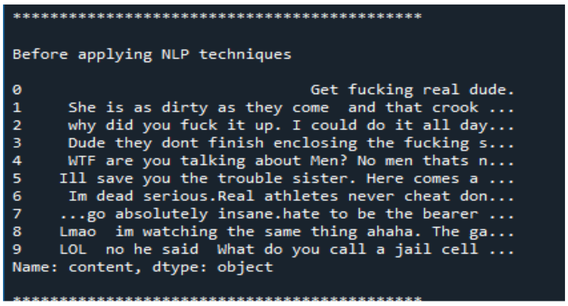
Figure 2: Pre-Processing of proposed system.

4.2 NLP Techniques :

Natural language pre-processing was a field in machine learning with the ability of a computer to analyze, understand, manipulated and potentially generate human language. The data pre-processing contains various steps as follows;

- ***Remove Punctuation***: Punctuation could provide grammatical context to a sentence that supports system for understanding.
- **Tokenization*:*** The texts were separated into units like words (or) sentences. It provides structure to previously unstructured text.
- ***Stemming:*** It supports to reduce a word to its stem form i.e., it lowers the inflection in words to their root term.

Sentiment evaluation: This process, evaluates the sentiment into positive, neutral and by utilizing sentiment analyzer, the negative terms are analyzed (polarity score). Sentiment analysis works through breaking a message down into chunks of topic and then sentiment score was assigned to each topic



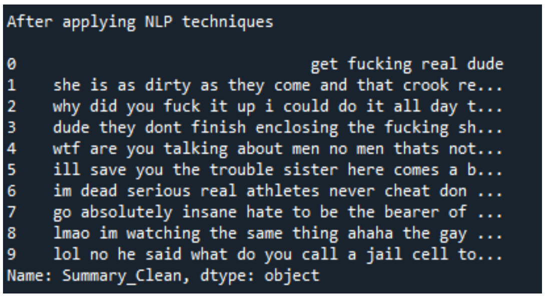
```
*******************************************

Before applying NLP techniques

0                                Get fucking real dude.
1     She is as dirty as they come  and that crook ...
2     why did you fuck it up. I could do it all day...
3     Dude they dont finish enclosing the fucking s...
4     WTF are you talking about Men? No men thats n...
5    Ill save you the trouble sister. Here comes a ...
6     Im dead serious.Real athletes never cheat don...
7    ...go absolutely insane.hate to be the bearer ...
8    Lmao  im watching the same thing ahaha. The ga...
9    LOL  no he said  What do you call a jail cell ...
Name: content, dtype: object

*******************************************
```

Figure 3: Pre-processing results before applying NLP techniques



```
After applying NLP techniques

0                                get fucking real dude
1    she is as dirty as they come and that crook re...
2    why did you fuck it up i could do it all day t...
3    dude they dont finish enclosing the fucking sh...
4    wtf are you talking about men no men thats not...
5    ill save you the trouble sister here comes a b...
6    im dead serious real athletes never cheat don ...
7    go absolutely insane hate to be the bearer of ...
8    lmao im watching the same thing ahaha the gay ...
9    lol no he said what do you call a jail cell to...
Name: Summary_Clean, dtype: object
```

Figure 4: Pre-processing result after applying NLP techniques

The figure 3 and 4 denotes the pre-processing results obtained before and after implementing the NLP techniques. This states that the NLP implemented system showed better results. Through this, the pre-processing was enhanced

Feature Extraction:

The proposed system performed feature extraction with count vectorizer.

4.3 Count Vectorizer:

It was basically used to transform a given text into a vector in terms of frequency of each word, which occurs in the whole text. Count-Vectorizer generates a matrix in that each unique word was denoted in matrix column and the text samples from the documents was denoted in the matrix row. The each cell value was nothing but the word count in particular text sample. This was helpful in terms of multiple such texts.

4.4 Decision Tree:

Decision tree was a flow chart like structure in that each nodes were denoted as "test" on attributed, test results were represented in each branch, and class label was denoted by each leaf node. The root to leaf path denotes the rules of classification. In this system decision tree was known as the proposed system. The algorithm for decision tree technique is shown in below:

```
Decision_Tree(Sample s, Features f)

Steps:

1.  If stopping condition(s, f) = True then

        a.  Leaf: createNode()
        b.  leaf label = classifies(s)
        c.  return_leaf

2.  root = createNode()

3.  root test condition = findBestSpilI( s,f )

4.  W = {w | w a probable outcome froot test
    condition]
```

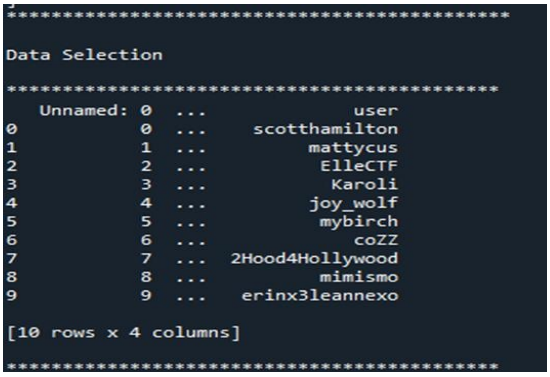
```
5.    For each measure v belongsto V:
      a.    S = {s root test condition(s) = v and S fits
            to s}:
      b.    Child = TreeGrowth(s, f);
      c.    child is added as descent of label and root
            the edge {root —>child}as v
6.    return_root
```

5. IMPLEMENTATION AND RESULTS

In this section, the dataset used for evaluation was explained and then the performance metrics measured for the evaluation was explained & the over-all performance study was explained. Followed by this, the evaluation of existing system was stated and then the proposed method evaluation was explained.

5.1 Dataset Description:

Since the system detects cyberbullying in social media, the proposed system used tweets collected by Twitter API streaming with the support of around 32 cyberbullying keywords such as kill, ban, hate, die, black, racism, swear, insult and threat[14][15][16]. The dataset tweets includes many outliers. Only English language tweets are considered and the other language tweets were removed and the retweets are filtered. Figure 5 shows the result of data selection.



```
**********************************************

Data Selection

*********************************************
   Unnamed: 0  ...              user
0           0  ...    scotthamilton
1           1  ...         mattycus
2           2  ...          ElleCTF
3           3  ...           Karoli
4           4  ...        joy_wolf
5           5  ...         mybirch
6           6  ...             coZZ
7           7  ...  2Hood4Hollywood
8           8  ...         mimismo
9           9  ...   erinx3leannexo

[10 rows x 4 columns]

*********************************************
```

Figure 5. Data selection result

5.2 Performance Metrics:

The efficiency of the system was determined by the performance metrics such as precision, recall, accuracy and F1-score.

1) Accuracy :

The word accuracy signified the attained intactprecisesorting of the measured dataset. Therefore, it was stated as the below-shown equation.

$$Accuracy = \frac{(True\_Positive + True\_neagtive)}{True\ negative + True\ postive + False\ neagtive + False\ negative}$$

2) Recall :

It was precise as the segment of retrieved-text and the relevant text of the data set to the relevant-text, it was articulated in the below equation.

$$Recall = \frac{relevant\_text \cap retrieved\_text}{relevant\_text}$$

3) Precision:

$$Precision = \frac{TP}{TP + FP}$$

The precision was measured with the below shown equation.

FP = False positive TP = True positive

4) F1 score :

The F1- score was specified as the harmonic-mean rate of the recall and precision.

$$F1\ score = \frac{2 * (precision * recall)}{precision + recall}$$

5.3 Performance Analysis

According to the above stated enactment metrics the evaluation was planned system was done. The performance of existing and proposed method was stated on the following section,

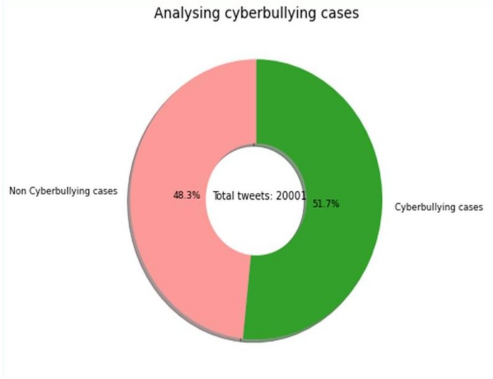


Figure 6: Analyzing cyberbullying cases

The figure 6 represents the overall analyzing of the cyberbullying cases. The twitter dataset were used for the detecting of cyberbullying in social media. Total 20001 tweets were selected for the analysis in that 48.3% were detected as non-cyberbullying cases and 51.7% were detected as cyberbullying cases.

5.4 Existing Method

The existing method considered in the proposed system was Naïve Bayes. Figure 5.3 represents the output attained in the Naïve Bayes method on all considered performance metrics

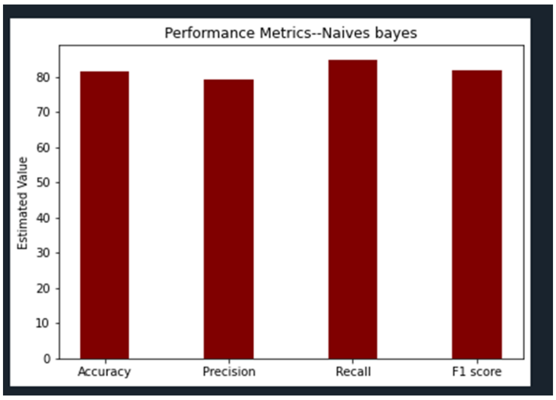


Figure 7: Performance analysis of the existing method –Naïve Bayes

The graphical representation of the existing Naïve Bayes system plotted with respect to the performance metrics and estimates values. The Naïve Bayes system showed 81.5% of accuracy, 79.1% of precision, 84.8% of recall and 81.8% of F1-score.Followed by this, the result obtained in the proposed method i.e., the result of decision tree with respect to all the performance were stated on the following section.

5.5 Proposed Method:

The proposed method utilized decision tree for the detection of cyberbullying in social media. The below shown figure 8 deliberates the obtained results of the proposed method.

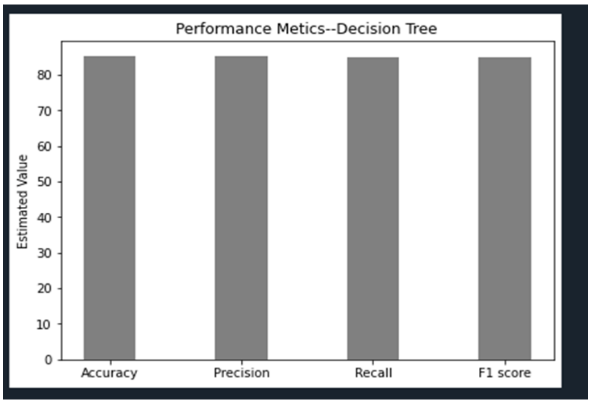
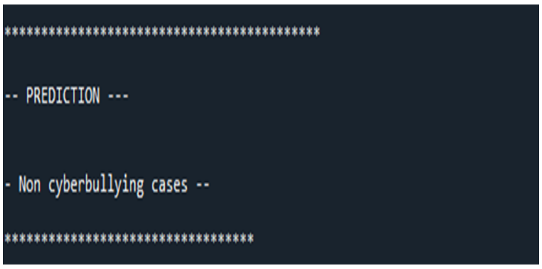


Figure 8: Performance analysis of the proposed system –Decision tree

The graphical representation of the proposed system was plotted with respect to the performance metrics and estimated values. The proposed system used decision tree for the cyberbullying detection. It showed 85.3% of accuracy, 85.3% of precision, 84.8% of recall and 85.0% of F1-score.Followed by this, the prediction result of the planned system was shown in figure 9.



```
*******************************************

-- PREDICTION ---


- Non cyberbullying cases --

**********************************
```

Figure 9: Prediction results

6. COMPARATIVE ANALYSIS

This section clearly denotes the comparative analysis, which helps to determine the competence and effectiveness of the planned system. The comparative investigation was accomplished with the existing method that is the considered Naïve Bayes techniques and the proposed Decision tree technique.

6.1 Comparative Analysis:

The comparative analysis was achieved to conclude the proficiency and effectiveness of the planned system. The system used decision tree and Naïve Bayes, using these two techniques the comparative analysis was performed.

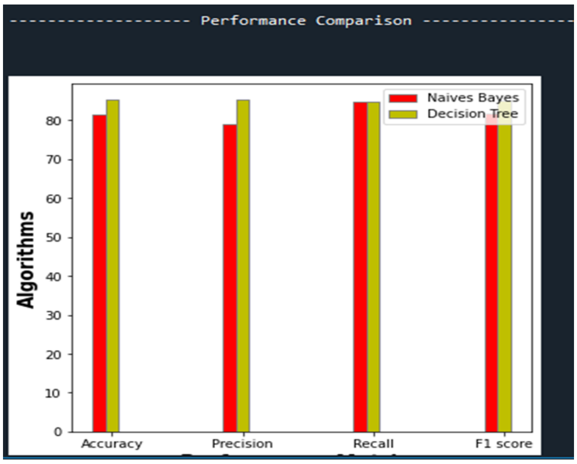


Figure 10: Comparative analysis result

7. CONCLUSION AND FUTURE WORK

Since social media was the platforms were the modern generation spends their time. It may contain both wanted and unwanted posts and message for teenagers (or) adolescents. Therefore, the system developed a mechanism for detecting cyberbullying activities in social media. If the proposed system was implemented successfully in detection of such post that are not appropriate for teenagers (or) adolescents, then the system can effectively deal with the crimes that were committed on the social media platforms. The approach was proposed for preventing and detecting twitter cyber-bullying using supervised binary classification ML algorithms. The prototypical was assessed on both the Naïve Bayes and decision tree, the feature extraction was performed with TFIDF vectorizer and the pre-processing was performed with NLP techniques and the classification was done with Decision tree. The decision tree showed 85.3% of accuracy, 85.3% of precision, 84.8% of recall and 85.0 % of F1-score were the obtained results are better than the existing Naïve Bayes methods. This proves that the proposed system was implemented successfully and showed its efficiency and effectiveness in the detection of cyberbullying.

In future, it was conceivable to modification (or) extension to the projected classification and clustering pseudo code to attain further amplified enactment. Apart from the experimental combination of data-mining methods, additional combinations with the other clustering algorithms could be used for improving the detection efficiency along with the reduction of offensive tweets. Hence, the cyberbullying detection approach could be stretched as an anticipation system to improve the system enactment.

REFERENCES:

[1] A. Ali and A. M. Syed, “Cyberbullying detection using machine learning,” *Pakistan Journal of Engineering and Technology,* vol. 3, pp. 45-50, 2020.

[2] M. Dadvar and K. Eckert, “Cyberbullying detection in social networks using deep learning based models,” in *International Conference on Big Data Analytics and Knowledge Discovery*, 2020, pp. 245-255.

[3] R. R. Dalvi, S. B. Chavan, and A. Halbe, “Detecting A twitter cyberbullying using machine learning,” in *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2020, pp. 297-301.

[4] A. Muneer and S. M. Fati, “A comparative analysis of machine learning techniques for cyberbullying detection on Twitter,” *Future Internet,* vol. 12, p. 187, 2020.

[5] A. Mangaonkar, R. Pawar, N. S. Chowdhury, and R. R. Raje, “Enhancing collaborative detection of cyberbullying behavior in Twitter data,” *Cluster Computing,* pp. 1-15, 2022.

[6] A. Sandesh, H. Asha, and P. Supriya, “Detection of Cyberbullying on Twitter Data Using Machine Learning,” in *Emerging Research in Computing, Information, Communication and Applications*, ed: Springer, 2022, pp. 703-713.

[7] J. Zhang, T. Otomo, L. Li, and S. Nakajima, “Cyberbullying detection on twitter using multiple textual features,” in *2019 IEEE 10th International Conference on Awareness Science and Technology (iCAST)*, 2019, pp. 1-6.

[8] V. Balakrishnan, S. Khan, and H. R. Arabnia, “Improving cyberbullying detection using Twitter users’ psychological features and machine learning,” *Computers & Security,* vol. 90, p. 101710, 2020.

[9] M. F. López-Vizcaíno, F. J. Nóvoa, V. Carneiro, and F. Cacheda, “Early detection of cyberbullying on social media networks,” *Future Generation Computer Systems,* vol. 118, pp. 219-229, 2021.

[10] Y. Chen, Y. Zhou, S. Zhu, and H. Xu, “Detecting offensive language in social media to protect adolescent online safety,” in *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Conference on Social Computing*, 2012, pp. 71-80.

[11] K. Jedrzejewski and M. Morzy, “Opinion mining and social networks: A promising match,” in *2011 International Conference on Advances in Social Networks Analysis and Mining*, 2011, pp. 599-604.

[12] H. Hosseinmardi, S. A. Mattson, R. Ibn Rafiq, R. Han, Q. Lv, and S. Mishra, “Analyzing labeled cyberbullying incidents on the instagram social network,” in *International conference on social informatics*, 2015, pp. 49-66.

[13] K. Reynolds, A. Kontostathis, and L. Edwards, “Using machine learning to detect cyberbullying,” in *2011 10th International Conference on Machine learning and applications and workshops*, 2011, pp. 241-244.

[14] K. Pareek, A. Choudhary, A. Tripathi, and K. Mishra, “Comparative Analysis of Social Media Hate Detection over Code Mixed Hindi-English Language,” in *Advances in Data and Information Sciences*, ed: Springer, 2022, pp. 551-561.

[15] A. Saravanaraj, J. Sheeba, and S. P. Devaneyan, “Automatic detection of cyberbullying from twitter,” *IRACST-International J. Comput. Sci. Inf. Technol. Secur,* vol. 6, pp. 2249-9555, 2016.

[16] A. Bozyiğit, S. Utku, and E. Nasibov, “Cyberbullying detection: Utilizing social media features,” *Expert Systems with Applications,* vol. 179, p. 115001, 2021.