

# CYBERBULLYING DETECTION USING NAIVE BAYES ALGORITHM

Muskaan Tabasum<sup>1</sup>, Fouzia Sultana<sup>2</sup>, Sumayya Maheen Unnisa<sup>3</sup> and Dr. Pathan Ahmed Khan<sup>4</sup>

<sup>1,2,3</sup>B.E STUDENT, <sup>4</sup>ASSISTANT PROFESSOR

*Department of CSE, ISL Engineering College, Bandlaguda, Hyderabad.*

**ABSTRACT** - In the era of social media and networking, the usage of bad words and aggressive words has been increased significantly. The young population is playing a major role in it. Cyberbullying affects more than half of the young population using social media. Insults in social media websites create negative interactions within the network. These remarks build up a culture of disrespect in cyberspace. Algorithms and tools used to understand and mitigate it are mostly inactive. Also, current implementations on insult detection using machine learning and natural language processing have very low recall rates. In short, the paper involves determining ways to identify bullying in text by analyzing and experimenting with different methods to find the feasible way of classifying such comments. We proposed an efficient algorithm to identify the bullying text and aggressive comments and analyse these comments to check the validity. Machine learning is used for analyzing the social comment and identified the aggressive effect of an individual or a group. An effective classifier acts as the core component in a final prototype system that can detect cyberbullying on social media.

## I. INTRODUCTION

Usage of unwanted words or cyberbullying has been increased on cyber platforms by the young community. It is increasing day by day due to active participation on social media. Users are increasing in an exponential manner and their time devotion is also increasing in the same manner. They are free to say anything on these platforms without any rules and regulation. It affects the accessibility of other people for these cyber or social platforms.

The use of social media has grown exponentially over time with the growth of the Internet and has become the most influential networking platform in the 21st century. However, the enhancement of social connectivity often creates negative impacts on society that contribute to a couple of bad phenomena such as online abuse, harassment, cyberbullying, cybercrime and online trolling. Cyberbullying frequently leads to

serious mental and physical distress, particularly for women and children, and even sometimes force them to attempt suicide. Online harassment attracts attention due to its strong negative social impact. Many incidents have recently occurred worldwide due to online harassment, such as sharing private chats, rumours, and sexual remarks. Therefore, the identification of bullying text or message on social media has gained a growing amount of attention among researchers.

## II. LITERATURE SURVEY

### 2.1 A Study of Cyberbullying Detection Using Machine Learning Techniques

Authors: Saloni Mahesh Kargutkar; Vidya Chitre

Abstract: Cyberbullying disturbs harassment online, with alarming implications. It exists in different ways, and is in textual format in most social networks. There is no question that over 1.96 billion of them would have an inescapable social operation. However, the developing decade presents genuine difficulties and the online-conduct of clients have been put to address. Expanding instances of provocation and harassing alongside instances of casualty has been a difficult issue. Programmed discovery of such episodes requires smart frameworks. A large portion of the current studies have been moving towards this issue with standard machine learning models and most of the models produced in these studies are scalable at one time into a solitary social network. Deep learning based models have discovered ways in the identification of digital harassing occurrences, asserting that they can beat the restrictions of the ordinary models, and improve the discovery execution. However, numerous old-school models are accessible to control the incident, the need to successfully order the tormenting is as yet weak. To successfully screen the harassing in the virtual space and to stop the savage outcome with the execution of Machine learning and Language preparing. A system is proposed to give a double characterization of cyberbullying. Our technique utilizes an inventive idea of CNN for content examination anyway the current strategies utilize a guileless way to deal with furnish the

arrangement with less precision. A current dataset is utilized for experimentation and our system is proposed with other existing methods and is found to give better precision and grouping.

## 2.2 Cyberbullying Detection on Social Networks Using Machine Learning Approaches

Authors: Md Manowarul Islam; Md Ashraf Uddin; Linta Islam; Arnisha Akter; Selina Sharmin.

Abstract: The use of social media has grown exponentially over time with the growth of the Internet and has become the most influential networking platform in the 21st century. However, the enhancement of social connectivity often creates negative impacts on society that contribute to a couple of bad phenomena such as online abuse, harassment cyberbullying, cybercrime and online trolling. Cyberbullying frequently leads to serious mental and physical distress, particularly for women and children, and even sometimes force them to attempt suicide. Online harassment attracts attention due to its strong negative social impact. Many incidents have recently occurred worldwide due to online harassment, such as sharing private chats, rumours, and sexual remarks. Therefore, the identification of bullying text or message on social media has gained a growing amount of attention among researchers. The purpose of this research is to design and develop an effective technique to detect online abusive and bullying messages by merging natural language processing and machine learning

## 2.3 Cyberbullying detection and prevention: Data mining and psychological perspective

Authors: Sourabh Parime; Vaibhav Suri

Abstract: Bullying is defined as targeting an individual or a group of individuals and exposing them to ridicule and negative actions both physical and mental deliberately. This is a common but serious and demoralizing experience that every individual encounters at least once in his or her lifetime. With the advent of technology, a form of bullying known as cyberbullying has spread very quickly targeting masses of innocent people very easily. Cyberbullying involves the use of computers, mobile phones, etc. for bullying activities. In this paper we focus on the data mining and machine learning techniques which have been proposed to detect and prevent cyberbullying and implement one such machine learning technique to identify the presence or absence of cyberbullying using the dataset from a popular social networking website. We also discuss the psychological factors related to cyberbullying and how the problem can be tackled along those factors. A few proposals for the future

algorithms for the detection and prevention of cyberbullying are also put forth.

## III. PROBLEM STATEMENT

Bullying is defined as targeting an individual or a group of individuals and exposing them to ridicule and negative actions both physical and mental deliberately. This is a common but serious and demoralizing experience that every individual encounters at least once in his or her lifetime. With the advent of technology, a form of bullying known as cyberbullying has spread very quickly targeting masses of innocent people very easily. Cyberbullying involves the use of computers, mobile phones, etc.

Cyberbullying is a term used for the content or images placed on online platforms which is not socially accepted. Abusing words, Aggressive words used by a group and individual. Twitter, YouTube, Instagram and many more are such kind of platforms where people used some unsocial words and give their opinion in very rude or aggressive manner. This is also a kind of online harassment. In USA this is considered as online threat. The major problems in fighting cyberbullying include: finding these kind of words and sentences when it occurred on online platforms; put forward to these cases in Law Agencies and finding the responsible persons. No present online community or social media websites (for example, Facebook and Twitter; where cyberbullying is most common), incorporates a system to automatically and intelligently identify aggression and instances of online harassment on its platform. Due to non-seriousness of this major issue earlier it is not considered the issue of research but now it is in dangerous phase. No-one can ignore this effect on cyber platform. It require a serious attention by researchers and cyber crime agencies to control this activity.

## IV. PROPOSED SYSTEM

We tried to develop a real time cyberbullying system which will identify bullying and non-bullying words. we are developing a prototype that can automatically detect cyberbullying and abusive behavior on social media and online communities. new words get added to slang regularly, we update the words using webscraping for dedecting the new slang abusive, harassment and bullying words.

### Advantages of proposed system:

In proposed System, we are using web scraping techniques to scrape the headlines and include them into

variables to improve our insights. This way our system won't get outdated soon. We are implementing an Email feature so that users can directly contact to helpline agencies for help chat application development.

## V. CONCLUSION

In this project work, we create ways to identify bullying in text by analyzing and experimenting with different methods to find the feasible way of classifying such comments. We proposed a efficient algorithm to identify the bullying test and aggressive comments and analyses these comments to check the validity. this work is to combat online harassment and aggression by developing a prototype that can automatically detect cyberbullying and abusive behavior on social media and online communities by Extracting, collecting, and labelling the data set and Preprocessing, cleaning, and experiment with various features to improve accuracy Classification of text, comment, or posts into one of the many classes Evaluation and analysis of best model. Machine learning is used for analyzing the social comment and identified the aggressive effect of an individual or a group. An effective classifier acts as the core component in a final prototype system that can detect cyberbullying on social media.

## VI. REFERENCES

- [1] H. Hosseinmardi, S. A. Mattson, R. I. Rafiq, R. Han, Q. Lv, and S. Mishra. Analyzing Labelled Cyberbullying Incidents on the Instagram Social Network. In In SocInfo, 2015.
- [2] Y. Chen, Y. Zhou, S. Zhu, and H. Xu. Detecting Offensive Language in Social Media to Protect Adolescent Online Safety. In PASSAT and SocialCom, 2012.
- [3] G. E. Hine, J. Onalapo, E. De Cristofaro, N. Kourtellis, I. Leontiadis, R. Samaras, G. Stringhini, and J. Blackburn. A Measurement Study of 4Chan's Politically Incorrect Forum and its effort on the web. In ICWSM, 2017.
- [4] N. Djuric, J. Zhou, R. Morris, M. Grbovic, V. Radosavljevic, and N. Bhamidipati. Hate Speech Detection with Comment Embeddings. In WWW, 2015.
- [5] I. Kayes, N. Kourtellis, D. Quercia, A. Iamnitchi, and F. Bonchi. The Social World of Content Abuser in Community Question Answering. In WWW, 2015.
- [6] C. Nobata, J. Tetreault, A. Thomas, Y. Mehdad, and Y. Chang. Abusive Language Detection in Online User Content. In WWW, 2016.
- [7] K. Dinakar, R. Reichart, and H. Liebrman. Modelling the Detection of Textual Cyberbullying. The Social Mobile Web, 11, 2011.
- [8] C. Van Hee, E. Lefever, B. Verhoeven, J. Mennes, B. Desmet, G. sDe Pauw, W. Daelemans, and V. Hoste. Automatic Detection and Prevention of Cyberbullying. In Human and Social Analytics, 2015.
- [9] V. Nahar, S. Unakard, X. Li, and C. Pang. Sentiment Analysis for Effective Detection of Cyberbullying. In APWeb, 2012.
- [10] J. M. Xu, X. Zhu, and A. Bellmore. Fast Learning for Sentiment Analysis on Bullying. In WISDOM, 2012.