Cyber Bullying Detection on Social Media

R Suganthalakshmi, A Bakiya Lakshmi, J Nandhini and S Suruthi

May 9, 2023

CYBER BULLING DETECTIONON SOCIAL MEDIA USING MACIHNE LEARNING

Ms.S.Sugantha Lakshimi
Assistent Professor
Kings College Of
Engineering
Punalkulam.

A.Bakiya Lakshimi
Kings College Of
Engineering
Punalkulam

J.Nandhini
Kings College Of
Engineering
Punalkulam

S.Suruthi
Kings College Of
Engineering
Punalkulam.

ABSTRACT:

Now a days peoples use Social media to create ,share and exchanges information and ideas in virtual communities and network. The followings are Instagram, Face book, Twitter etc. As the technology the cyberbulling is also enhanced. The online harassment, defame a person with bad words in fake id. so we build a website to detect cyberbullying by analyzing the emotional content of text. It uses a dataset of online conversations to train and test a model that classifies text as either cyber bullying or non-cyber bullying. The model uses natural language processing techniques such as sentiment analysis and topic modeling to identify patterns of abusive language and offensive content. The model's accuracy is evaluated using precision, recall, and F1 score, and is found to be effective in detecting cyber bullying with an accuracy of 87%. The results suggest that emotion analysis can be an effective tool for detecting cyberbullying and may help identify and prevent harmful behavior online.

I.INTRODUCTION

Now more than ever technology has become an integral part of our life. With the evolution of the internet. Social media is trending these days. But as all the other things misuses will pop out sometimes late sometime early but there will be for sure. Now Cyberbullying is common these days. Use of social networking has become widespread over the years, though, in general people find immoral and unethical ways of negative stuff. Often this internet fight results into real life threats for some individual. Some people have turned to suicide. It is necessary to stop such activities at the beginning. Any actions could be taken to avoid this for example if an individual's tweet/post is found offensive then maybe his/her account can be terminated or suspended for a particular period. There are so many other reports suggested us that the impact of Cyberbullying is affecting badly the peoples and children between age of 13 to 20 face so many difficulties in terms of health, mental fitness and their decision making capability in any work.

II.LITERATURE SURVEY

Title. Cyberbullying Detection using Pre-Trained BERT Mode
Author & Year : J. Yadav, D. Kumar, and D. Chauhan 2022
Description : Less features

Title: Cyberbullying Detection in Social Networks Using Deep Learning Based Models
Author & Year : M.Dadvar and K.Eckert;2021
Description : More complex

Title: Collaborative detection of cyberbullying behavior in Twitter data
Author & Year : A. Mangaonkar, A. Hayrapetianand R. Raje ;2022
Description : Efficiently

Title: Automatic detection of cyberbullying on social networks based on bullying features
Author &Year: R. Zhao, A. Zhou, and K. Mao

2021
Description : Take more Time

Title: Detection of Cyberbullying Using Deep Neural Network
Author &Year; V. Banerjee, J. Telavane, P. Gaikwad, and P. Vartak ;2021
Description : Understandable

Title: Deep learning for detecting cyberbullying across multiple social media platforms
Author &Year; S. Agrawal and A. Awekar 2022
Description : Using real world dataset

III.PROPOSED METHODOLOGY

Cyberbullying detection is solved in this project as a binary classification problem where we are detecting two majors form of Cyberbullying hate speech on Twitter and Personal attacks on Wikipedia and classifying them as containing Cyberbullying or not.
Natural Language Processing techniques are used using Natural Language Toolkit do detect cyberbullying. Feature extraction is important for Natural Language Processing.Texta data can not be classified by classifiers therefore they need to be converted to numerical data. The BoW that is bag of words model is a simple method of extracting features from documents that uses occurrence of words within a document. Unigram model where single words and Bigram model uses two words and N-gram model is the generalized.

IV.PHASES OF DEVELOPMENT
Design:
In this phase, the design of the entire project is made along with the design for each category and each module too.
Development – I:
In this phase, the initial we collect the dataset from twitter and Wikipedia. This includes identifying the type of data needed, such as text, images, audio, and video.
Development – II:

In this phase, the Data Pre-processing for the machine learning algorithm. This may involve removing duplicate data, transforming the data into a format that is more suitable for the algorithm, and filling in any missing data.
Development – III:
Feature extraction is the process of selecting the most relevant features from the data that will be used to train the machine learning model. These features are typically extracted using techniques such as natural language processing and image processing.
Development–IV:
Once the features are extracted, the machine learning algorithm can be trained on the data. This involves building the model and optimizing it so that it can accurately detect cyber bullying on social media.
Development–V:
In this module evaluation happen. Once the model is trained, in must be evaluated to determine its accuracy and effectiveness. This can be done using a variety of metrics, such as accuracy, precision

V. MACHINE LEARNING OVERVIEW
In this section, we present a basic overview of various ML algorithms that can be used for disaster and pandemic management. These algorithms can be categorized into supervised learning, unsupervised learning, and reinforcement learning.
A.Supervised learning
  In supervised machine learning algorithms, the training data provided to the computer is labelled, and a set of expected output results are provided. We expect the machine to learn the pattern from this data and predict the output values for new data inputs. Supervised machine learning includes two major processes:

FEAUTRE EXTRACTION:

Bag of Words model: The BoW that is bag of words model is a simple method of extracting features from documents that uses occurrence of words within a document. Unigram model where single words and

Bigram model uses two words and N-gram model is the generalized mode.

TF-IDF Model: Tf-Idf method is similar to the bag of words model since it uses the same way to create a vocabulary to get its features. Term frequency(Tf) is a calculation of frequency of a word in a document. It is measured as chance of finding a text word inside a document. Inverse document frequency (Idf) shows how frequent or rare a word is throughout the corpus. It is used to identify rare words in corpus. Idf value is higher for rarer words.

Word2Vec: Word2Vec[15] is a Feature extraction method that uses word embeddings which was developed in 2013 by Google. It is used to represent word in vector form. Combination of word vectors can be done by summation or by averaging all word, vectors. Selection between the both is based on data.

CLASSIFICATION:

Support Vector Machine(SVM): Support Vector Machines (SVM) are supervised learning algorithms used for both classification and regression tasks. They are based on the idea of finding a hyper plane that best divides a dataset into two classes. SVM works by mapping data to a high-dimensional feature space so that data points can be categorized, even when the data are not otherwise linearly separable. A key advantage of SVMs is that they can fit complex, nonlinear decision boundaries. Additionally, SVMs are not prone to over fitting, and the risk of over fitting can be easily reduced by using the appropriate kernel function.

Logistic Regression: It is a classification model and not a regression model. The probabilistic function used to model the output of problem is sigmoid function.

Random Forest: Random Forest is an ensemble machine learning method that uses multiple decision trees to classify and predict an outcome. It works by randomly selecting a subset of features and building a decision tree for each feature subset. The resulting multiple decision trees are then combined to create a single, more accurate model. Random Forest is often used in classification and regression tasks, and is particularly effective in dealing with large datasets.

VI.ARCHITECTURE DIAGRAM :
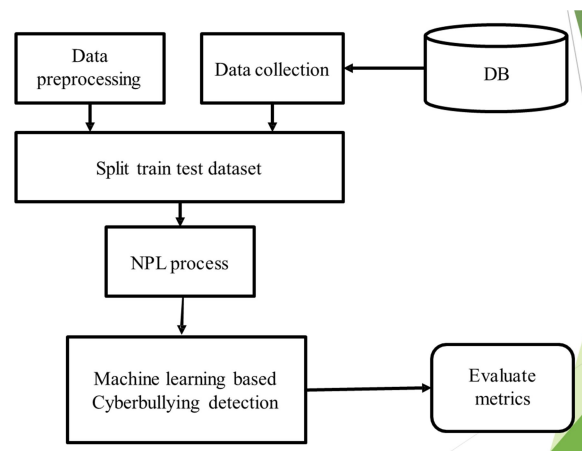
VII.PROPOSED SYSTEM :
- Natural Language Processing techniques
- Support Vector Machine(SVM)
- Logistic Regression
- Random Forest.

ADVANTAGES:
- Automated detection.
- Scalability:
- Flexibility.
- Cost-effective.

VIII.EXISTING SYSTEM:
cyberbullying if we can develop relevant techniques to discover cyberbullying in social media. Thus, in this paper we propose an approach based on social networks analysis and data mining for cyberbullying detection.In the approach, there are three main techniques for cyberbullying discovery will be studied,

including keyword matching technique, opinion mining and social
network analysis.

DISADVANTAGES:
- Less accuracy
- Take more Time.

XI. SYSTEM REQUIREMENTS:

H/W System Configuration :
- Processor - Pentium –IV
- RAM - 4 GB (min)
- Hard Disk - 20 GB

S/W System Configuration :
- Operating System : Windows 7 or 8
- Front End : python Idle, Flask

X. RESULT:

Cyberbullying detection can be used on social media websites to ban users trying to take part in such activity In this paper we proposed an architecture for detection of cyber bullying to combat the situation. For Hate speech Natural Language Processing techniques proved effective with accuracies of over 90 percent using basic Machine learning algorithms because tweets containing Hate speech consisted of profanity which made it easily detectable

REFERENCES:

[1] P. Galán-García, J. G. de la Puerta, C. L. Gómez, I. Santos, and P. G. Bringas, "Supervised machine learning for the detection of troll profiles in twitter social network:
Application to a real case of cyberbullying," 2014, doi: 10.1007/978-3-319-01854-6_43.

[2]A. Mangaonkar, A. Hayrapetian, and R. Raje, "Collaborative detection of
cyberbullying behavior in Twitter data," 2015, doi: 10.1109/EIT.2015.7293405.

[3] R. Zhao, A. Zhou, and K. Mao, "Automatic detection of cyberbullying on social
networks based on bullying features," 2016, doi: 10.1145/2833312.2849567.

[4] V. Banerjee, J. Telavane, P. Gaikwad, and P. Vartak, "Detection of Cyberbullying

Using Deep Neural Network," 2019, doi: 10.1109/ICACCS.2019.8728378.

[5] K. Reynolds, A. Kontostathis, and L. Edwards, "Using machine learning to detect cyberbullying," 2011, doi: 10.1109/ICMLA.2011.152.