# CYBER OFFENSIVE PREDICTIVE ANALYSIS ON TWITTER DATA USING MACHINE LEARNING

**Sanjay R 1 , Gautam R 2 , Mrs.Jayashankari J 3 , Dr.Preetha M 4**

**Student, Information Technology, Prince Shri Venkateshwara Engineering College 1,2**

**Assistant Professor, Department of Information Technology 3**

**Professor, Department of Computer Science and Engineering 4**

**Prince Shri Venakteshwara Padmavathy Engineering College**

## Abstract:

Cyber-bullying refers to the utilization of aggressive, violent, or offensive language, targeting a selected cluster of individuals sharing a typical property, whether or not thisproperty is their gender, their grouping or race, or their beliefs and faith. With the risingof social networks, communication between folks from completely different cultural and psychological backgrounds has become a lot direct, leading to a lot of ‘‘cyber’’ conflictsbetween these folks and it's become a significant downside. Therefore, arises the need to discover such speech mechanically in the Twitter victimization Machine Learning rule. The system depends on the detection of 3 basic tongue elements equivalent to Insults, Swears, and person. Here is the classification of, the real-world dataset from Twitter, one in every of the highest 5 networks with the highest share of cyber-bullying instances. An arrangement named Random Forest, call tree and logistical regression has been utilized to discover the incidence of cyber entities in Twitter. These algorithms are an effective method of Decision Making because they: Clearly lay out the problem so that all options can be challenged, allow us to analyze fully the possible consequences of a decision, and provide a framework to quantify the values of outcomes and the probabilities of achieving them. The call tree classifier can be used in both classification and regression. It can help represent the decision as well as make a decision. The Random Forest classifier is consisting of multiple decision tree classifiers. Each tree gives a class prediction individually. The maximum number of the predicted class is our final result. The projected system prediction model is to use a text classification approach that involves the development of machine learning classifiers from labeled text instances. Datasets containing bullying texts, messages orposts are collected and prepared for the machine learning algorithms using natural language processing. In our projected model, at first information preprocessing is completed and so tokenization and normalization can happen. The processed datasets are then used to train the machine learning algorithms for detecting any harassing or bullying message on social media including Facebook and Twitter. The potency and accuracy are high within the projected model due to multi-model algorithms specifically random forest, call tree and logistical regression. Additionally, this model may acknowledge all sorts of matter inputs and predict the output.

**Keywords:** Regression algorithm, Naïve Bayes Algorithm, Decision tree Algorithm, Django, Flask

## 1.Introduction:

Cyberbullying or cyber-harassment is a form of bullying or harassment using electronic means. Cyberbullying and cyber-harassment are also known as online bullying. It has become increasingly common, especially among teenagers, as the digital sphere has expanded and technology has advanced. Cyberbullying is when someone, typically a teenager, bullies or harasses others on the internet and in other digital spaces, particularly on social media sites. Harmful bullying behavior can include posting rumors, threats, a victims personal information, or pejorative labels (i.e. hate speech). Bullying or harassment can be identified by repeated behavior and an intent to harm. Victims of cyber-bulling may experience lower self-esteem, increased suicidal ideation, and a variety of negative emotional responses including being scared, frustrated, angry, or depressed.

A social network is a social structure made up of a set of social actors (such as individuals or organizations), sets of dyadic ties, and other social interactions between actors. The social network perspective provides a set of methods for analyzing the structure of whole social entities as well as a variety of theories explaining the patterns observed in these structures. The study of these structures uses social network analysis to identify local and global patterns, locate influential entities, and examine network dynamics. The social network is a theoretical construct useful in the social sciences to study relationships between individuals, groups, organizations, or even entire societies (social units, see differentiation). The term is used to describe a social structure determined by such interactions. The ties through which any given social unit connects representthe convergence of the various social contacts of that unit. This theoretical approach is, necessarily, 12 relational. An axiom of the social network approach to understanding social interaction is that social phenomena should be primarily conceived and investigated through the properties of relations between and within units, instead of the properties of these units themselves.

## 2.Literature survey:

**[1]** Tadashi Nakano et.al., proposed to develop software tools that help users of social networking sites intervene cyber-bullying incidents at early stages of cyber-bullying or and preventcyber-bullying incidents before they may occur. This paper focuses on an online social networkingsite Ask.fm, withthe goal of understanding user behavior that may potentially lead to cyber- bullying incidents on social networking sites. This paper considers Ask.fm, a social networking site where users create profiles and can send each other questions and analyses aggressive user behavior that may potentially lead to cyber-bullying incidents. They hypothesize that anonymity isa primary cause of such aggressive user behavior and examine how anonymous and non- anonymous users behave in social networking. They collected data from Ask.fm and analyzed questions posted by anonymous and non-anonymous users and answers posted bynonanonymous users. Analysis of the collected data shows that anonymous users exhibit more aggressive behavior than non-anonymous users. Analysis also shows that users become more aggressive in answering aggressive anonymous questions than aggressive non-anonymous questions.

**[2]** Hajime Watanabe et.al., proposed an approach to detect hate expressions on Twitter. The rapid growth of social networks and microblogging websites, communication between peoplefrom different 16 cultural and psychological backgrounds has become more direct, resulting in more and more ‘‘cyber’’ conflicts between these people. Consequently, hate speech is used moreand more, to the point where it has become a serious problem invading these open spaces. Hatespeech refers to the use of aggressive, violent or offensive language, targeting a specific group of people sharing a common property, whether this property is their gender, their ethnic group orrace (i.e., racism) or their believes and religion. While most of the online social networks and micro blogging websites forbid the use of hate speech, the size of these networks and websites makes it almost impossible to control all of their content. Therefore, arises the necessity to detectsuch speech automatically and filter any content that presents hateful language or language inciting to hatred.

This approach is based on unigrams and patterns that are automatically collected from thetraining set. These patterns and unigrams are later used, among others, as features to train a machine learning algorithm. Our experiments on a test set composed of 2010 tweets show that our approach reaches an accuracy equal to 87.4% on detecting whether a tweet is offensive or not (binary classification), and accuracy equal to 78.4% on detecting whether a tweet is hateful, offensive, or clean (ternary classification).

**[3]** Charles Lim et.al. aimed to create a text classification system that classifies the document using several algorithms. The rise of computer violence, such as Distributed Denial ofService (DDoS), web vandalism, and cyberbullying are becoming more serious issues when theyare politically motivated and intentionally conducted to generate fear in society. These 17 kinds of activity are categorized as cyber terrorism. As the number of such cases increases, the availability of information regarding these actions is required to facilitate experts in investigating cyber terrorism. This research aims to create a text classification system that classifies the document using several algorithms including Naïve Bayes, Nearest Neighbor, Support Vector Machine (SVM), Decision Tree, and Multilayer Perceptron. The result shows that SVM outperforms by achieving 100% of accuracy. This result concludes the excellent performance of SVM in handling high-dimensional of data.

**[4]** Kelly Reynolds et.al. proposed to detect language patterns used by bullies and their victims and develop rules to automatically detect cyberbullying content. Cyberbullying is the use of technology as a medium to bully someone. Although it has been an issue for many years, the recognition of its impact on young people has recently increased. Social networking sites provide a fertile medium for bullies, and teens and young adults who use these sites are vulnerable to attacks. Through machine learning, we can detect language patterns used by bullies and their victims, and develop rules to automatically detect cyberbullying content. The data they used for their project was collected from the website Formspring. me, a question-and-answer formatted website that contains a high percentage of bullying content. The data was labeled using a web service, Amazon's Mechanical Turk. We used the labeled data, in conjunction with machine learning techniques provided by the Weka tool kit, to train a computer to recognize bullying content. Both a C4.5 decision tree learner and an instance-based learner were able to identify the true positives with 78.5% accuracy.

**[5]** Yee Jang Foong proposed an online system for automatic detection and monitoring of Cyber-bullying cases from online forums and online communities. Cyber-bullying has recently been reported as one that causes tremendous damage to society and economy. Advances in technology related to web-document annotation and the multiplicity of the online communities renders the detection and monitoring of such cases rather difficult and very challenging. This

paper describes an online system for automatic detection and monitoring of Cyber-bullying cases from online forums and online communities. The system relies on the detection of three basic natural language components corresponding to Insults, Swears, and Second Person. A classification system and ontology-like reasoning have been employed to detect the occurrence of such entities in the forum/web documents, which would trigger a message to security in order to take appropriate action. The system has been tested on two distinct forums and achieves reasonable detection performances.

**[6]** Arturo Elias et.al. focused on the social and cultural implications of cyber technologies. Identity, bullying, and inappropriate use of communication are major issues that need to be addressed in relation to communication technologies for security in web use. The contribution ofthis paper is topresent a novel approach to explain the performance of a novel Cyberbullying model applied on a Social Network using multi-agents to improve the understanding of this socialbehavior.

**[7]** Efthymios Lalas et.al. proposed architecture for easing the aforementioned problems and demonstrates it through its implementation on Facebook. Furthermore, the main advantages of the proposed approach are discussed in terms of future implementations on other social networking sites.

**[8]** Mingmei Li Atsushi Tagami et.al. focused on detecting relation-based cyber-bullying, whichis an indirect attack on a human, e.g., isolating a victim by ignoring the victim’s messages. Recently, relationship-based cyber-bullying has received attention as a new type of cyber-bullying, and detecting it is still a novel problem. As it attacks a human relationship, the detectionshould monitor the change of the human relationship. In this paper, for the first step of relation- based cyber- bullying detection, we propose a framework to generate a contact network. The framework consists of two phases for a reduction of false negatives, i.e. students are friends in the school but detected as non-friend in the Social Networking Service (SNS), which is a seriousproblem for cyber-bullying detection. Finally, this paper analyzes the collected SNS data with theactual human relationships and evaluates the proposed framework. 20

**[9]** Nektaria Potha et.al., proposed the study of the accuracy of predicting the level of cyberbullying attacks using classification methods and also to examine potential patterns betweenthe linguistic style of each predator. More specifically, unlike previous approaches that consider a fixed window of a cyber predators questions within a dialogue, they exploit the whole question set and model it as a signal, whose magnitude depends on the degree of bullying content. Using

feature weighting and dimensionality reduction techniques, each signal is straightforwardlyparsed by a neural network that forecasts the level of insult within a question given a window between two and three previous questions. Throughout the time series modeling experiments, aninteresting discovery was made. By applying SVD to the time series data and taking into accountthe second dimension, we observed that its plot was very similar to the plot of the class attribute.By applying a Dynamic Time Warping algorithm, the similarity of the aforementioned signals wasproved to exist, providing an immediate indicator for the severity of cyberbullying within a given dialogue.

**[10]** Fujio Toriumi et.al. proposed the communication behaviors by focusing on private chat systems, which are sometimes used to sexually entrap minors. Our analysis found that most of the users communicate with less than 5users, and the communication ended within a half hour. Then they classified users into 15 clusters based on their communication behaviors. From the clustering results, we identified both active communication senders and active communication receivers.

**[11]** Asaf Varol et.al., proposed experimentally succeeded to attack an Android smartphone using a simple software-based radio circuit. they have developed a software “Primary Mobile Hack Builder” to control Android-operated cellphones at a distance. The SMS information and pictures on the cellphone can be obtained using this device. On the other hand, after launching software targeting cellphones, the camera of the cellphone can be controlled for taking pictures and downloading them to our computers. It was also possible to eavesdropping the conversation.

**[12]** Daphney–Stavroula Zois et.al., demonstrated the effectiveness of their approach using areal-world dataset from Twitter, one of the top five networks with the highest percentage of users reporting cyberbullying instances. We show that our approach is highly scalable while not sacrificing accuracy for scalability.

**[13]**Siz Chen et.al., build a Long Short-Term Memory Neural Network Deterministic Finite Automaton (LND) model which considers not only the language content but also the user's characteristics and historical speech on social networks. Due to the lack of labeled content, we utilize Douban's 22 reviewers' data by analyzing speech patterns with polarized emotions. Then the learned model is applied to analyze Chinese cyberbully behaviors on Weibo.

**[14]** Mukul Anand et.al., classified the comments in the following categories: toxic, severe toxic, obscene, threat, insult, and identity hate. The dataset is trained with various deep learning

## 2 3.Existing System:

3 The existing system uses three basic Natural Language Processing: Insult, Swear and Second Person.

4 They used Support Vector Machine algorithm for classifying the cyberbullying text.

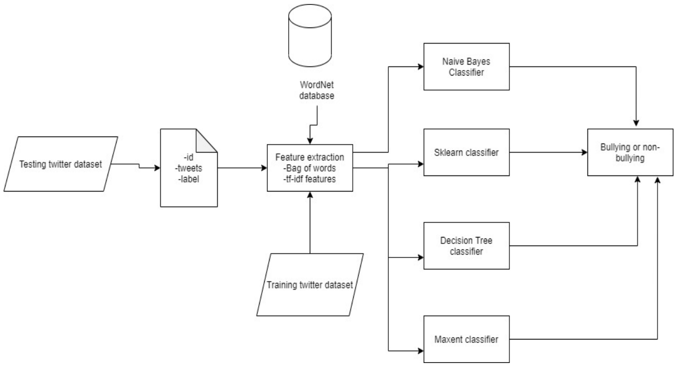
5 They used the social media platform ASK.fm where users ask questions publicly on other users pages. It also provides possibility to ask questions anonymously as well as possibility to view samples of users profiles.

6 The Scrapy crawler has been used for this purpose. They deliberately select questions / answers that contain at least one word that belong to Insult or Swear category using a simple string matching function.

7 They separately store the usernames of the questioners. An SQL server database was employed in order to store and index all the dataset attributes (usernames, questions, posts, date, links), which ultimately boost the indexing and retrieving functions. A total of around 10,000 questions and answers were gathered from the site.

8 Around 17% of the total collected dataset is found to entail genuine cyberbullying cases, after an initial brief scrutiny of the dataset. The entire data set was then split into a training set and a testset.70 percent of both the negative and positive examples were used as a training set while the remaining 30 percent were used as the test set Prior to subsequent analysis, automatic pre-processing procedure assembles the comments for each user and chunks them into sentences.

## 9 4.System Architecture:



**5.Algorithm Used:**

There are 2 types of algorithms used they are: 1.Naive bayes classifier

2.Decision Tree classifier

### *NAÏVE BAYES CLASSIFIER*

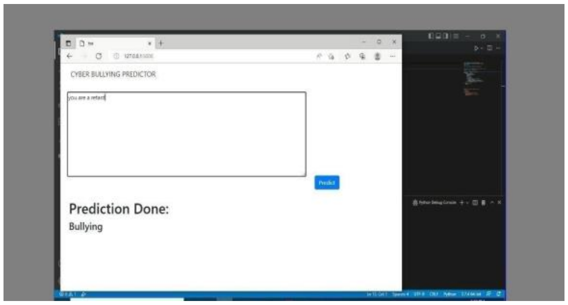
Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayestheorem and used for solving classification problems. It is mainly used in text classification that includes a high-dimensional training dataset. Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions. It is a probabilistic classifier, which means it predicts on the basis of the probability of an object. Some popular examples of Naïve Bayes Algorithm are spam filtration, Sentimental analysis, and classifying articles.

### *DECISION TREE CLASSIFIER*

Decision Tree is a supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.

In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches. The decisions or the test are performed on the basis of features of the given dataset. It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions. It is called a decisiontree because, similar to a tree, it starts with the root node, which expands on further branches and constructs a tree-like structure. In order to build a tree, we use the CART algorithm, which stands for Classification and Regression Tree algorithm. A decision tree simply asks a question, and based on the answer (Yes/No), it further split the tree into subtrees.

## 6.Result And Conclusions



CYBER BULLYING PREDICTOR

you are a retard

Prediction Done:

Bullying

Thus an effective model is built and the machine is well trained in order to achieve high accuracy. A classification using ML algorithms have been employed to detect the occurrence of cyber-bullying entities in Twitter, which would identify whether the given text contains bullying or non-bullying. A combination of features have been employed to get better performances , combinations include Bag-of-Words and TF-IDF features. The proposal uses natural language processing techniques and machine learning in order to infer whether a post belongs to bullying category or not. A combination of features have been employed. This includes standard TF-IDF whose weights are boosted for those terms that belong to Insult / Swear category and Bag-of- Words feature . The experiment also investigated the performance of the classifier when using various combination of the aforementioned features.

The proposal uses natural language process. Thus an effective model is built and the machine is well trained in order to achieve high accuracy. A classification using ML algorithms have been employed to detect the occurrence of cyber- bullying entities in Twitter, which would identify whether the given text contains bullying or non- bullying.

## REFERENCES:

**[1] Amirita Dewani , Mohsin Ali Memon, "Cyberbullying detection: advanced preprocessing techniques & deep learning architecture for Roman Urdu data" ,journal of Big data, Springer,2021**

**[2] Muhammad Owais Raza, Mohsin Memon, Sania Bhatti, Rahim Bux, "Detecting Cyberbullying in Social Commentary Using Supervised Machine Learning, Advances in Intelligent Systems and Computing book series (AISC, volume 1130), Feb 2020**

**[3] Md Manowarul Islam, Md Ashraf Uddin, Linta Islam, Arnisha Akter, Selina Sharmin, "Cyberbullying Detection on Social Networks Using Machine Learning Approaches", 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), 2020**

**[4] N Novalita, A Herdiani, I Lukmana, D Puspandari, "Cyberbullying identification on Twitter using random forest classifier", The 2nd International Conference on Data and Information Science, 2019**

**[5] Kelvin Kiema Kiilu, George Okeyo, Richard Rimiru and Kennedy Ogada, "Using Naive Bayes Algorithm in detection of Hate Tweets", International Journal of Scientific and Research Publications, vol. 8, no. 3, March 2018.**

**[6] Kirti Kumari, Jyoti Prakash Singh, Yogesh Kumar Dwivedi, Nripendra Pratap Rana, "Towards Cyberbullying-free social media in smart cities: a unified multi-modal approach", Soft Computing volume 24, pages11059–11070, November 2019**

**[7] Akshi Kumar, Shashwat Nayak, Navya Chandra, "Empirical Analysis of Supervised Machine Learning Technique**