Summary:

due to the huge amount of data transferred daily through the social media platforms it is impossible to use the humans in detecting the cyberbully. Cyberbully is a problem that is facing a lot of teens and could lead to a crisis like in sometimes committing suicide and to overcome these problems we are going to talk about various ways that had been proposed by various people and then we are going to present out proposed model that help the people to detect the cyberbully over the social media by all of its sorts like (Facebook , Instagram , Whats app , Twitter ,….etc).

2- Proposed Method:

According to Vikas S Chavan and Shylaja S S in their paper “Machine Learning Approach for Detection of Cyber-Aggressive Comments by Peers on Social Media Network” they introduced their method which have very good results in detecting the cyberbully but the accuracy wasn’t very good.

2.1- System Overview:

2.1.1- Pre-processing

2.1.2-Standard Feature extraction

2.1.3-Additional Features

2.1.4-Feature Selection

2.1.5-Classification

2.2-Explaining their proposed method.

They are going to pre-process the raw comments with various ways to enhance and make the feature extraction and classification very efficient. Their pre-processing consists of 2 main parts the “removing unwanted strings” and “correcting words”. The removing unwanted strings part , they remove the unwanted strings like the punctuation and the encoding parts. But , for the correcting words parts they tend to correct the misspelled words and also the words that has been written wrong to by pass the classification system like the word “Idiot” any one could write it “!d!iot” for the humans it remains readable but for the classifier it is a different word from the word it has trained on from the dataset.

Second part the standard feature extraction. They used they feature extraction methods for this part. First , N-gram method, they used the n-gram method to extract the tokens they are going to work on , they used the unigram , bigram , trigram . Second , counting , in this part they are going to tokenize the comments and count the occurrence of every token . Third , TF-IDF , they used it to calculate the importance of the words according to their number of occurrence in the document.

Third part is the additional features. They used 2 methods in this part which are capturing pronouns and skip-gram. First , capturing pronouns , they used this method to help capturing the cyberbully by the occurrence of the pronoun “You” in the comment. Second , skip-gram , they used this method to detect the co-occurrence of some words.

Fourth part is feature selection. For this part they used the chi- square which is a method that is used to select the best features to work on as the classifier couldn’t use all the extracted features so the chi – square helps solving this problem by the chi – square formula that detects the best features to work on.

Last part , classification. They used 2 classifiers for this which are the logistic regression and support vector machine. The logistic regression achieved 73.76 accuracy and 0.6 recall and 0.644 Precision . But for the support vector machine they achieved 77.65 accuracy and 0.58 recall and 0.7 precision.

3- Datasets:

they collected datasets from kaggle

4-Drawbacks:

in the pre-processing part they used to correct the words that is written to by pass their classifier according to their paper they are going to use the dictionary to make this work but if the model encountered a new word that isn’t available in the dictionary it is going to by pass it without alerting the user that there is a cyberbully in this comment. Also they haven’t used any sentiment or contextual analysis to make sense of the words which is going to make a big problem of false positive because their could be 2 friends talking and use any word that is in the dictionary of the bullying words so this is going to raise alert flag without any actual bullying.